

КОНЦЕПТУАЛЬНОЕ ПРОЕКТИРОВАНИЕ ИНФОРМАЦИОННОЙ СИСТЕМЫ ПРОГНОЗИРОВАНИЯ ДЕМОГРАФИЧЕСКИХ ПОКАЗАТЕЛЕЙ РЕГИОНА

Будаев Е.С.

Восточно-Сибирский государственный университет технологий и управления
E-mail: budeus@mail.ru

Развитие информационных технологий диктует создание информационных систем в различных отраслях экономики. Создание демографической системы, как и создание любой предметной системы имеет свою специфику, которая связана со способами добычи, хранения и обработки демографической информации. В статье рассматриваются основные моменты концептуального проектирования информационной системы прогнозирования демографических показателей региона.

Ключевые слова: Информационная система, проектирование информационных систем, прогнозирование данных, демография, региональная информационная система.

DESIGNING AN INFORMATION SYSTEM FOR FORECASTING DEMOGRAPHIC INDICATORS OF THE REGION

E.S. Budaev

East Siberian State University of Technology and Management
E-mail: budeus@mail.ru

The development of information technology dictates the creation of information systems in various sectors of the economy. The creation of a demographic system, like the creation of any subject system, has its own specifics, which is associated with the methods of extracting, storing and processing demographic information. The article discusses the main points of designing an information system for predicting the demographic indicators of a region.

Key words: Information system, information systems design, data forecasting, demography, regional information system.

Концептуальная модель интеллектуальной информационной системы демографических прогнозов региона

Тенденции развития современных экономических информационных систем показывает, что эффективная система должна строиться с учетом постоянно меняющихся данных, с использованием современных технологий обработки больших данных, использования методов искусственного интеллекта. В таком случае, необходимо создавать интеллектуальную информационную систему. Такие системы, помимо традиционного использования данных государственных реестров населения, могут использовать и данные цифрового следа отдельных граждан – на основе использования больших данных из социальных сетей, данных операторов сотовой связи, операторов интернет-доступа и т.д.

Работа с данными из различных источников с одной стороны требует специальных подходов для безошибочного слияния данных, с другой стороны, дает большие возможности для аккумуляции различной информации о населении для проведения анализа демографических процессов и выявления их зависимостей от других социально-экономических показателей. Поэтому создание единой интеллектуальной информационной системы, обладающей достоверностью и системностью демографических данных является актуальной задачей. Она позволит, на макроуровне выполнять как краткосрочное, так и долгосрочное прогнозирование, нацеленное на повышение уровня жизни населения региона,

а на локальном уровне – лицам, принимающим решения на различных уровнях управления, эффективно и в кратчайшие сроки оказывать государственные услуги.

Выявление, изучение и прогнозирование региональных особенностей демографического процесса является одной из важнейших предпосылок стабильного функционирования социально-экономической системы страны в целом на долгосрочную перспективу. Поэтому в работе предлагается проектирование региональной информационной системы моделирования и прогнозирования демографических показателей, учитывающей, во-первых, не только демографические закономерности социально-экономического развития страны, но и региональные особенности, влияющие на демографическую ситуацию региона, во-вторых, нацеленной на создание благоприятных условий для решения социально-экономических проблем региона.

Результаты проведенных исследований послужили теоретической основой разработки механизма принятия решений в процессе прогнозирования демографической ситуации региона в виде единой распределенной интеллектуальной системы (рис. 1), способной адаптироваться под изменения внешней среды и пополнять свои знания с учетом современных научных исследований и потребностей в области демографического прогнозирования.

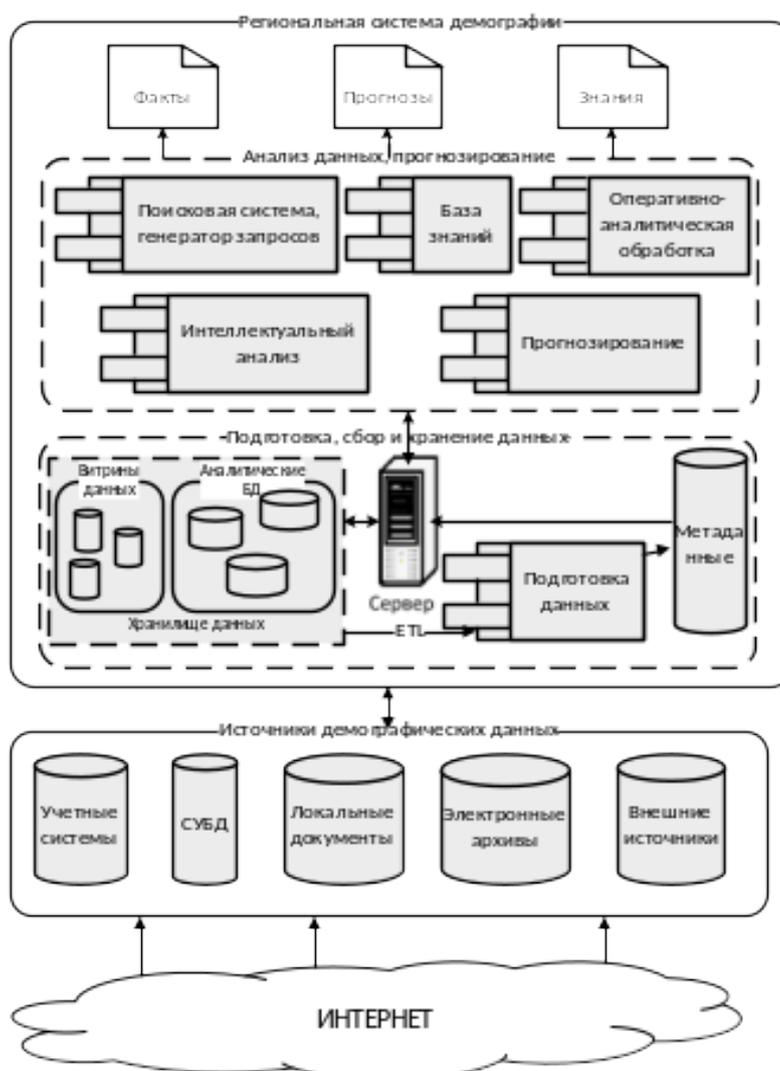


Рисунок 1 – Концептуальная модель региональной информационной системы демографии

Для решения задач, возложенных на информационную систему, предусматривается программная интеграция с существующими информационными системами и пакетами прикладных программ:

- единый регистр населения;
- информационные системы государственных министерств, департаментов и ведомств;
- пакеты статистических прикладных программ;
- инструменты для визуализации данных;
- географические информационные системы.

Предлагаемая система преследует следующие основные цели:

- систематизация, структуризация и интеграция методологических основ и используемого программного обеспечения в анализе и прогнозировании численности населения;
- формирование единых сквозных процессов прогнозирования численности населения верхнего уровня на основе процессов, имеющих в разных подсистемах государственного, регионального и муниципального управления, а также формирование интеллектуального центра принятия решений по выявлению и решению демографического кризиса;
- формализация и интеллектуализация процессов демографического прогноза на основе методов искусственного интеллекта, в том числе с применением базы знаний;
- обеспечение вариантности процессов демографического прогнозирования путем интеллектуального анализа прогнозного фона в условиях высокого динамизма внешней среды;
- обеспечение непрерывности прогнозирования за счет повышения скорости обработки и предоставления статистической информации;
- обеспечение корректности работы системы на основе верификация моделей и сценариев прогнозирования демографических показателей в режиме реального времени;
- сокращение трудозатрат расчета прогнозных демографических показателей, необходимых государственным органам при разработке программ социально-экономического развития на определенный период.

В процессе разработки системы подлежат реализации следующие основные процессы:

- цифровизация демографических показателей и методологии прогнозирования;
- цифровизация сценариев демографического прогнозирования;
- сбор и подготовка данных, оценка полноты собранной информации для прогнозирования;
- оценка и прогноз демографических показателей;
- интеллектуальный анализ данных;
- верификация моделей и сценариев прогнозирования;
- визуализация статистических данных;
- выработка альтернативных рекомендаций для лица, принимающего решение, по принятию комплекса мер для достижения требуемого прогнозного демографического показателя.

В интеллектуальной информационной системе заложены два подхода к реализации демографического прогноза:

1) *классическое прогнозирование* – это моделирование демографической динамики посредством существующих информационных моделей и выбор наилучшей из них;

2) *прогнозирование на основе методов искусственного интеллекта* – это моделирование демографической динамики посредством искусственной нейронной сети, с применением механизма обучения с учителем, тестирования нейронной сети, переобучения, с последующей валидацией полученного прогноза с результатами классического прогнозирования.

Второй подход предоставляет технологическую основу для проведения следующих видов исследований:

- прогноз целевого демографического показателя без изменения факторов, влияющих на данный показатель;
- прогноз целевого демографического показателя при условии изменения факторов, влияющих на данный показатель;
- решение обратной задачи: какими должны быть факторы, чтобы достичь требуемого целевого показателя.

Обобщая выше сказанное, основные технологические процессы прогнозирования демографических показателей информационной системе можно представить в виде обобщенной схемы (рис.2).

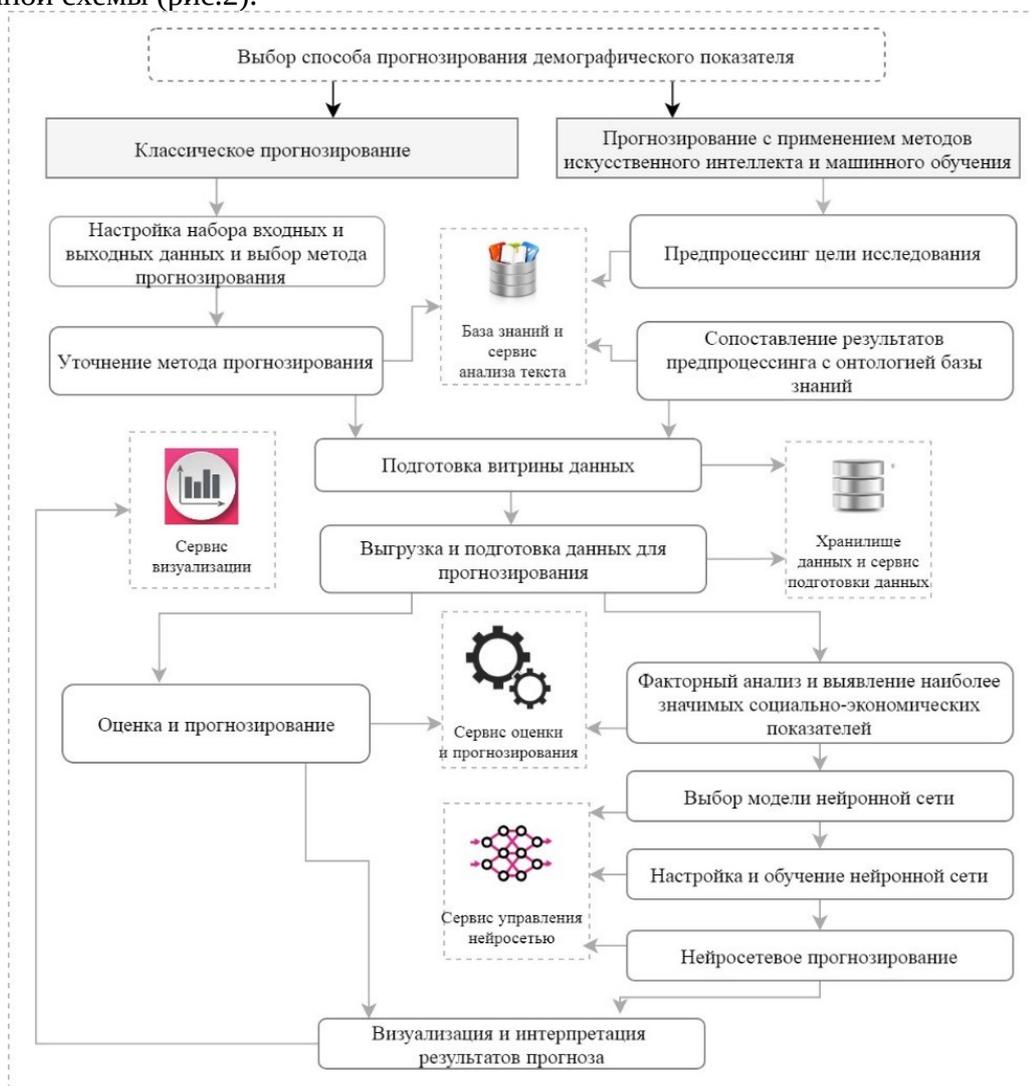


Рисунок 2 – Обобщенная схема технологических процессов в интеллектуальной информационной системе демографического прогноза региона

На данной схеме представлены модель взаимодействия технологических процессов интеллектуальной системы с базой знаний и сервисом анализа текста, сервисом визуализации, хранилищем данных и сервисом подготовки данных. Таким образом, создаваемая система позволит проводить прогноз демографических показателей как с помощью классических методов, так и посредством применения методов искусственного интеллекта.

Проектирование структуры интеллектуальной информационной системы демографических прогнозов региона

На основе выполненной концептуальной модели интеллектуальной информационной системы демографических прогнозов региона разработана обобщенная схема системы (рис. 3), которая включает в себя пять ее основных компонентов: база знаний о демографии, подсистема подготовки демографических данных, подсистема оценки и прогнозирования демографических показателей, хранилище данных о демографической ситуации, подсистема интеллектуального анализа и прогнозирования демографических показателей.

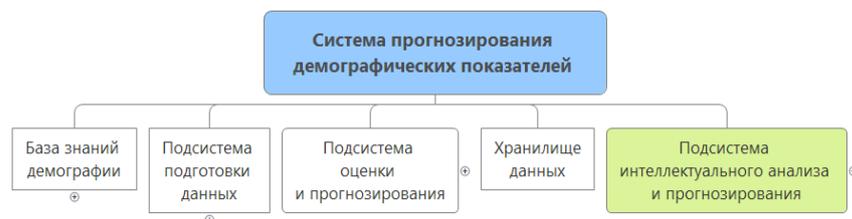


Рисунок 3 – Обобщенная схема интеллектуальной информационной системы демографических прогнозов региона

Рассмотрим более подробно, что представляют собой основные компоненты будущей интеллектуальной информационной системы:

1. **База знаний демографии** представляет собой открытую модель в виде онтологии предметной области, которая служит, во-первых, для представления знаний о демографических понятиях и их отношениях, и во-вторых, для управления этими знаниями в процессах прогнозирования демографических показателей.

2. **Подсистема подготовки данных** предназначен для решения задачи взаимодействия с внешними информационными системами и пакетами прикладных программ с целью подготовки исходных данных. Основными функциями подсистемы являются ввод, сбор или экспорт данных из внешних источников, анализ полноты данных, группировка и первичная обработка данных, их передача в хранилище данных.

3. **Подсистема оценки и прогнозирования** включает разные алгоритмы, механизмы и инструменты управления процессом моделирования, прогнозирования, интерпретации и визуализации прогнозируемых данных.

4. **Хранилище данных** предназначено для долговременного хранения данных о показателях демографических процессов. В хранилище должны храниться как исходные данные, на основе которых были осуществлены моделирование и прогнозирование, так и сами прогнозные данные.

5. **Подсистема интеллектуального анализа и прогнозирования**, основной задачей которой является многомерный анализ данных, включает анализ структуры и тенденций демографических показателей, выделение и анализ факторов, влияющих на демографические показатели. Основными функциональными элементами подсистемы являются модуль машинного обучения для формирования обучающей выборки и подготовки нейронной сети для дальнейшего прогнозирования, модуль многомерного анализа для проведения интеллектуального анализа и сравнения больших данных, выявления наиболее значимых факторов, влияющих на демографический показатель, модуль прогнозирования на обученной нейронной сети и модуль верификации моделей для автоматического сравнения прогнозных данных с реальными, автоматического выявления рассогласований и принятия решения о необходимости внесения изменений в информационные модели, сценарии, базу знаний. Результаты функционирования подсистемы в дальнейшем могут быть использованы для выработки комплекса мер, направленных на стабилизацию демографической ситуации в отдельном регионе.

На рис.4 представлена детальная схема интеллектуальной системы после дополнения вышеуказанной обобщенной схемы входящими в ее подсистемы модулями, который изображены на схемы на втором уровне декомпозиции.



Рисунок 4 – Схема интеллектуальной информационной системы

Детализированная схема будет основой для проектирования прототипа системы, на ней представлена навигация между формами будущей системы.

Проектирование интерфейса интеллектуальной информационной системы демографических прогнозов региона

Проектирование интерфейса информационной системы требует от разработчика полного понимания структуры, общего с командой разработки видения, того как будет работать системы, какие конкретно информационные технологии будут задействованы в разработке системы.

В случае концепция системы разработана, выбрана веб-технология в качестве среды разработки системы, сформирована детальная структура системы. В качестве технического средства разработки высокоуровневого прототипирования системы выбрано программное обеспечение Axige RP – один из лидеров средств проектирования интерфейса программ. Одним из удобных функций данного ПО является возможность оценки прототипа через онлайн-опубликование в облаке компании разработчика для дальнейшей совместной работы над интерфейсом. После выполненного прототипирования есть возможность представить готовый прототип системы в виде веб-страниц для встраивания на веб-узел разрабатываемого проекта.

На основе схемы интеллектуальной информационной системы (рис. 4) выполнено прототипирование и верификация перечисленных в концептуальной модели сервисов системы.

На рис.5 представлен макет главной страницы интеллектуальной информационной системы демографических прогнозов региона.



Рисунок 5 – Прототип главной страницы интеллектуальной информационной системы демографических прогнозов региона

Прототип модуля базы знаний (рис. 6) состоит из сервисов наполнения и редактирования базы знаний, формирования базы правил управления технологическими процессами планирования и прогнозирования, а также просмотра дерева понятий онтологии и отношений между ними.



Рисунок 6 – Прототип окна просмотра онтологии демографических показателей

Прототип подсистемы подготовки данных (рис. 7) включает в себя:
 – модуль взаимодействия с внешними системами, позволяющий осуществлять загрузку данных в хранилище из внешних источников;

- модуль сбора и подготовки данных, в котором реализованы алгоритмы проверки данных на полноту, а также производится процедура заполнения отсутствующих данных в витрине хранилища;
- модуль прогнозирования;
- модуль подготовки структуры хранилища в части его дополнения новыми показателями при расширении понятийного аппарата онтологии базы знаний.

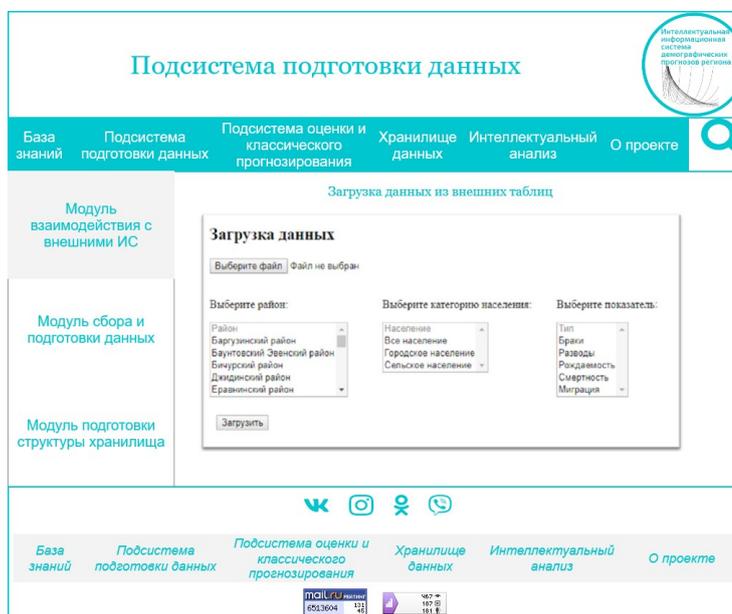


Рисунок 7 – Прототип подсистемы подготовки данных

- Прототип подсистемы оценки и классического прогнозирования включает в себя (рис. 8):
- конструктор моделирования демографических процессов, в котором задаются прогнозные целевые показатели и параметры входных данных, а также выбирается метод, с помощью которого необходимо осуществить прогнозирование;
 - модуль анализа и интерпретации данных, содержащий методы исследования взаимосвязей между демографическими показателями и социально-экономическими факторами;
 - модуль прогнозирования, который включает в себя программные методы краткосрочного прогнозирования, описанные выше;
 - модуль визуализации результатов анализа и прогнозирования.

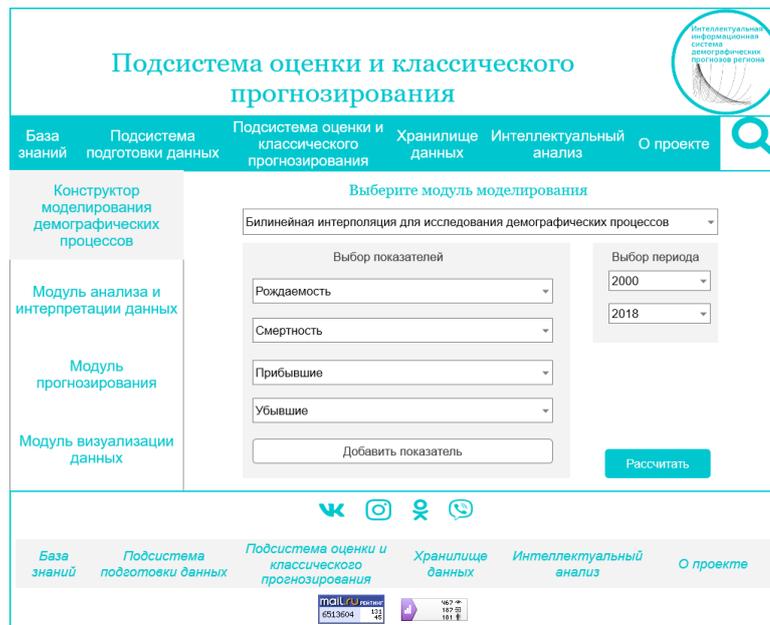


Рисунок 8 – Прототип подсистемы оценки и классического прогнозирования

Модуль анализа и интерпретации данных включает в себя программные алгоритмы двух методов анализа демографических данных: билинейной интерполяции (рис.9) и пространственного анализа на основе алгоритма Крускала.

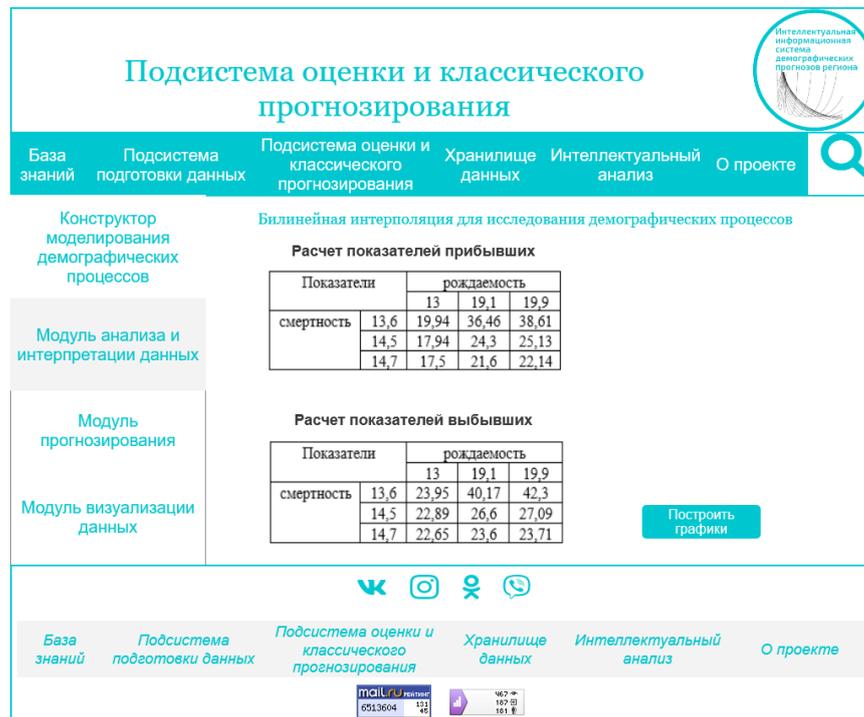


Рисунок 9 – Прототип модуля анализа данных билинейной интерполяции

В результате визуализации данных с помощью графиков можно не только проследить динамику относительных показателей демографических процессов, но и выдвинуть гипотезу о характере их зависимости для дальнейшего детального исследования. Результаты билинейной интерполяции (рис. 10) позволяют сделать вывод о присутствии некоторых закономерностей. Так, например, результаты расчета показателя выбывших по данным рождаемости и смертности схожи с исходными данными, а расчеты показателя прибывших имеют отличие в динамике изменения.



Рисунок 10 – Визуализация зависимостей расчетных и исходных значений показателей миграции с показателями рождаемости и смертности в виде графиков поверхностей

Алгоритм пространственного анализа демографических показателей на основе алгоритма Крускала. Алгоритм Крускала предназначен для построения минимального остовного дерева для содержательной характеристики региональной экономической динамики в контексте изменения макроэкономических параметров с учетом экономических прогнозов близлежащих регионов. На рис.11 представлена визуализация минимального остовного дерева для коэффициента младенческой смертности.

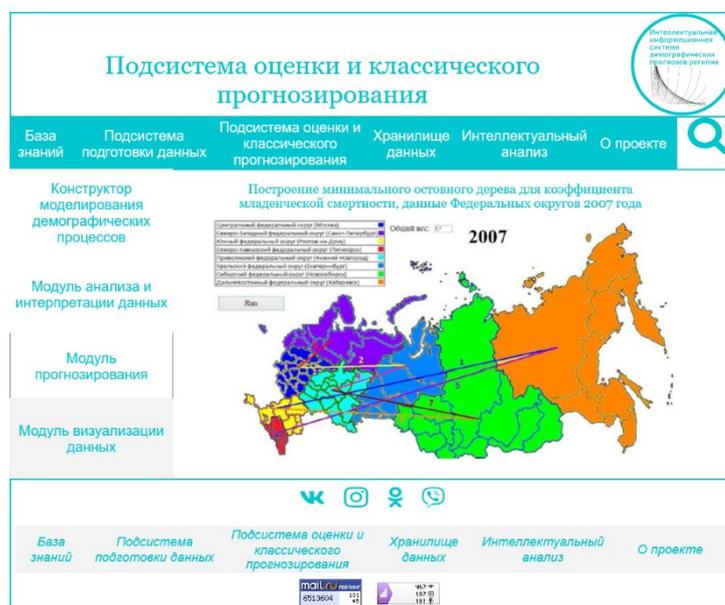


Рисунок 11 – Визуализация минимального остовного дерева для коэффициента младенческой смертности по данным Федеральных округов 2007 года

А рис.12 представлен прототип раздела работы с хранилищем данных. Здесь находятся разделы работы со списком баз данных и списком витрин данных.



Рисунок 12 – Макет раздела «Хранилище данных»

Прототип модуля интеллектуального анализа и нейросетевого прогнозирования (рис. 13) позволяет вводить цель на естественном языке, проводить предпроцессинг цели исследования, осуществляет наложение полученного результата на онтологию базы знаний (рис. 14), выполняет запросы в хранилище данных с целью получения исходных данных для машинного обучения, определяет структуру нейронной сети (рис. 15), осуществляет прогноз на нейронной сети по заданному временному периоду и отображает результаты прогнозирования в виде графика с сопоставлением расчетного значения с истинными значениями из хранилища данных (рис. 16).

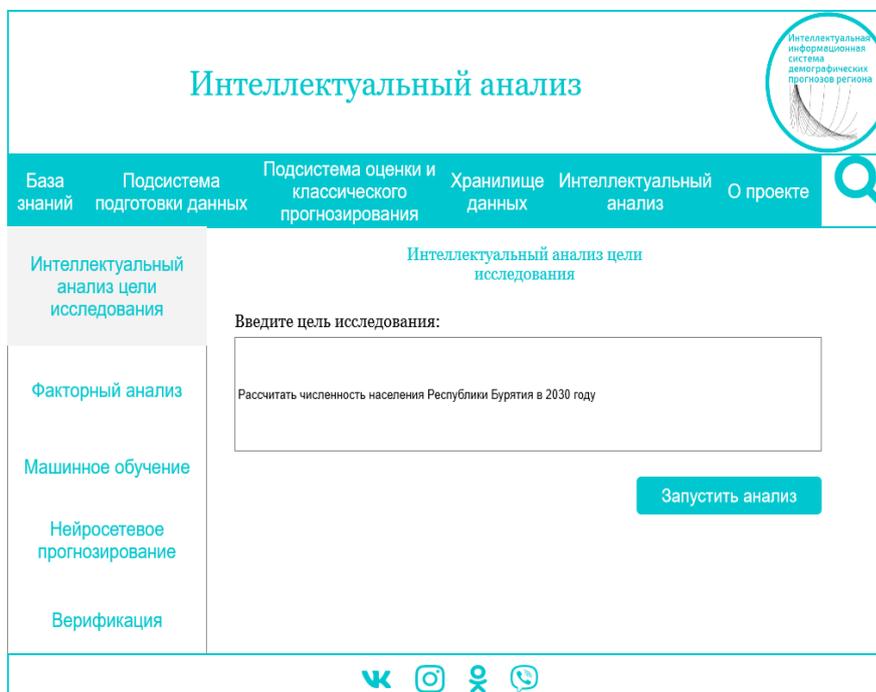


Рисунок 13 – Прототип интерфейса задания цели исследования на естественном языке



Рисунок 14 – Прототип визуализации результатов предпроцессинга и анализа на отнологии цели исследования

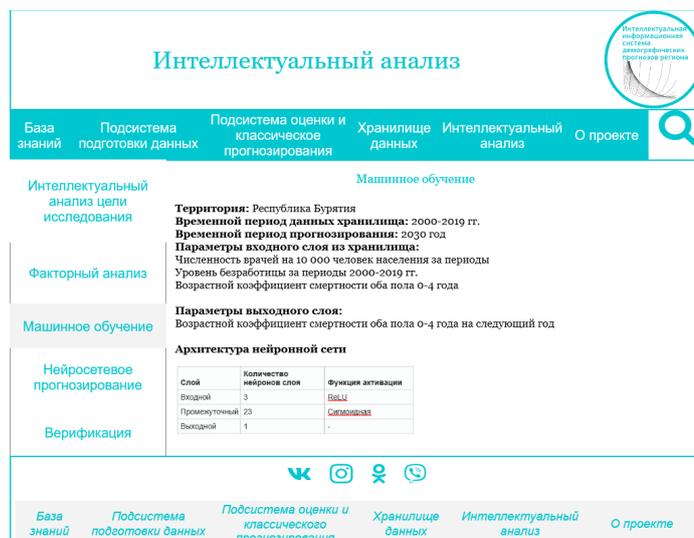


Рисунок 15 – Прототип визуализации результатов машинного обучения



Рисунок 16 – Прототип визуализации результатов нейропрогнозирования

Прототип веб-сервиса анализа демографических показателей региона (элемент результата работы представлен на рис. 17) входит в состав подсистемы оценки и прогнозирования. Созданный прототип описан в работах «Разработка модуля сайта анализа демографических показателей региона» [7] и «Development of the Web Service «Analysis of Demographic Indicators of the Region» [1].



Рисунок 17 – График населения Заиграевского района

Прототип веб-сервиса создан с использованием широкого круга различных инструментальных средств, начиная со средств разработки до использования специальных библиотек визуализации представленной информации в виде динамических карт, столбиковых, линейных и ступенчатых диаграмм. В ходе проекта был создан сайт, интерфейс которого разработан с учетом требований к эргономике пользовательского интерфейса.

В процессе работы над созданием проекта интеллектуальной информационной системы разработан прототип будущей системы в качестве макета высокой достоверности. Данный прототип информационной системы опубликован для ознакомления и находится по адресу: <https://bl2dpk.axshare.com>

Библиография

1. Budaev E.S., Mikhaylova S.S., Stepanov V.V., Development of the Web Service «Analysis of Demographic Indicators of the Region // 2nd International Scientific and

- Practical Conference on Digital Economy (ISCDE 2020), Yekaterinburg. Paris: Atlantis Press, 2020. Vol. 156. pp. 265-274
2. Будаев Е.С., Степанов В. В., Разработка веб-модуля динамической карты региона для анализа демографических показателей // Информационные технологии в экономике: материалы IV Республиканской научно-практической Интернет-конференции студентов, аспирантов и молодых ученых с международным участием, 25-26 марта 2020 года [Электронное издание]. – Луганск: Изд-во ЛНУ им. В. Даля, 2020. С. 137-141.
 3. Будаев Е.С., Экспорт базы данных смертности центра демографических исследований российской экономической школы // Развитие социально-экономических систем в условиях цифровизации: материалы Национальной научно-практической конференции ВСГУТУ «Образование и наука». – Улан-Удэ: Издательство ВСГУТУ, 2020, С. 33-39.
 4. Михайлова С. С., Шулунова И. Р., Евдокимова И. С., Будаев Е. С., Данилова С. Д. Методологические основы создания интеллектуальной информационной системы демографических прогнозов региона: монография. – Улан-Удэ: Изд-во ВСГУТУ, 2020. – 224 с.
 5. Степанов В. В., Димова А. В., Будаев Е.С., Шулунова И. Р., Разработка модуля сайта анализа демографических показателей региона // Приоритетные направления развития науки и образования: сборник статей VII VII Международной научно-практической конференции. В 2 ч. Ч. 1. – Пенза: МЦНС «Наука и Просвещение», 2019, С. 248

Bibliography

1. Budaev E.S., Mikhaylova S.S., Stepanov V.V., Development of the Web Service “Analysis of Demographic Indicators of the Region // 2nd International Scientific and Practical Conference on Digital Economy (ISCDE 2020), Yekaterinburg. Paris: Atlantis Press, 2020. Vol. 156. pp. 265-274
2. Budaev E.S., Stepanov V.V., Development of a web module for a dynamic map of the region for the analysis of demographic indicators // Information technologies in the economy: materials of the IV Republican scientific and practical Internet conference of students, graduate students and young scientists with international participation , March 25-26, 2020 [Electronic edition]. - Lugansk: Publishing house of LNU named after V. Dahl, 2020.S. 137-141.
3. Budaev ES, Export of the mortality database of the center for demographic research of the Russian economic school // Development of socio-economic systems in the context of digitalization: materials of the National Scientific and Practical Conference of VSGUTU "Education and Science". - Ulan-Ude: VSGUTU Publishing House, 2020, pp. 33-39.
4. Mikhailova S. S., Shulunova I. R., Evdokimova I. S., Budaev E. S., Danilova S. D. Methodological bases for creating an intellectual information system of demographic forecasts of the region: monograph. - Ulan-Ude: VSGUTU Publishing House, 2020 .-- 224 p.
5. Stepanov V.V., Dimova A.V., Budaev E.S., Shulunova I.R., Development of the site module for analyzing the demographic indicators of the region // Priority directions of development of science and education: collection of articles VII VII International scientific and practical conferences. At 2 pm Part 1. - Penza: ICNS "Science and Education", 2019, p. 248