

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение  
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»**

Кафедра дискретной математики и информационных технологий

**ПРОГНОЗИРОВАНИЕ ТЕНДЕНЦИЙ РАЗВИТИЯ IT-РЫНКА**

**АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ**

студентки 4 курса 421 группы  
направления 09.03.01 — Информатика и вычислительная техника  
факультета КНиИТ  
Соколовой Татьяны Владимировны

Научный руководитель  
профессор, д. э. н.

\_\_\_\_\_

Л. В. Кальянов

Заведующий кафедрой  
доцент, к. ф.-м. н.

\_\_\_\_\_

Л. Б. Тяпаев

Саратов 2023

## ВВЕДЕНИЕ

С момента, как для человечества наступила компьютерная эра, одной из самых важных областей информационных технологий стала «поддержка принятия решений». Сегодня эта сфера важна как никогда. Работая в динамичной и постоянно меняющейся среде, современные менеджеры несут ответственность за целый ряд далеко идущих решений: *«Должна ли компания увеличивать или сокращать штат сотрудников? Выйти на новые рынки? Разрабатывать новые продукты? Инвестировать в исследования и разработки?»* Список можно продолжить. Но, несмотря на присущую этим вопросам сложность и постоянно увеличивающийся объем информации, с которой приходится иметь дело бизнес-аналитикам, все эти решения сводятся к двум фундаментальным вопросам:

- Что может произойти в будущем?
- Какое лучшее решение на данный момент?

Осознаем ли это человечество или нет, но эти два вопроса пронизывают нашу повседневную жизнь — как на личном, так и на профессиональном уровне. Например, когда люди едут на работу, им нужно узнать прогнозирование пробок на дорогах, прежде чем выбрать самый быстрый маршрут. На работе нужно предсказать спрос на продукт, прежде чем решать, сколько его производить. И прежде чем инвестировать в зарубежный рынок, необходимо предсказать будущие обменные курсы и экономические переменные. Кажется, что независимо от принимаемого решения или его сложности сначала нужно сделать прогноз того, что может произойти в будущем, а затем принять наилучшее решение на основе этого прогноза. Этот фундаментальный процесс лежит в основе основной предпосылки *Adaptive Business Intelligence* — *адаптивной бизнес-аналитики*.

Адаптивная бизнес-аналитика — это дисциплина, объединяющая прогнозирование, оптимизацию и адаптируемость в систему, способную ответить на два фундаментальных вопроса: *«Что может произойти в будущем? Какое лучшее решение на данный момент?»* Чтобы построить такую систему, сначала нужно понять методы и приемы, обеспечивающие прогнозирование, оптимизацию и адаптируемость. Необходимо понять, как объединить бизнес-аналитику, методы интеллектуального анализа данных и прогнозирования, методы оптимизации в программную систему, способную прогнозировать, оптимизировать и адаптировать.

Будущее индустрии бизнес-аналитики тесно связано с системами, которые могут принимать решения, а не с инструментами, создающими подробные отчеты. Как теперь понимают большинство бизнес-менеджеров, есть огромная разница между хорошими знаниями и подробными отчетами и принятием разумных решений. Именно поэтому моя работа является актуальной на сегодняшний день.

Применимость адаптивной бизнес-аналитики для решения многих реальных бизнес-задач очень велика, начиная от прогнозирования спроса и планирования и заканчивая обнаружением мошенничества и инвестиционными стратегиями. Применение адаптивной бизнес-аналитики может обеспечить значительные преимущества и экономию.

*Целью* написания данной бакалаврской работы является разработка информационной системы для прогнозирования тенденций IT-рынка, используя язык программирования Python.

В связи с поставленной целью в данной работе решаются следующие задачи:

1. Рассмотреть специальную литературу по изучаемой проблеме.
2. Исследовать содержание ключевых понятий.
3. Изучить методы машинного обучения.
4. Изучить алгоритмы машинного обучения.
5. Проанализировать программные средства для построения и обучения нейронных сетей.
6. Реализовать прогнозирование с помощью нейронной сети.
7. Разработать пользовательский интерфейс.

Материалы исследования представляют собой литературу, посвященную бизнес-аналитике, прогнозированию и науке о данных.

В данной работе содержится восемь глав, а именно: «Адаптивная бизнес-аналитика», «Создание системы адаптивной бизнес-аналитики», «Наука о данных», «Интеллектуальный анализ данных и статистика для принятия решений», «Наделение компьютеров способностью обучаться на данных», «Нейронные сети», «Программные средства для построения и обучения нейронных сетей», «Программная реализация».

## КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

**В первой главе** рассматривается понятие *адаптивной бизнес-аналитики*.

Ранее бизнес-менеджеры хранили огромные объемы данных, полагая, что они содержат ценную информацию. Но в конце концов было обнаружено, что необработанные данные редко имеют какую-либо пользу, и что реальная ценность данных зависит от способности организации анализировать их. Следовательно, возникла потребность в программных системах, способных извлекать, обобщать и интерпретировать данные для конечных пользователей.

Эта потребность способствовала появлению сотен компаний, занимающихся *бизнес-аналитикой (business intelligence)*, которые специализируются на предоставлении программных систем и услуг для извлечения знаний из необработанных данных [1].

Сегодня большинство бизнес-менеджеров осознают, что существует разница между наличием необходимых знаний и принятием правильного решения. Поскольку этот пробел влияет на способность руководства отвечать на фундаментальные вопросы бизнеса (такие как «Что нужно сделать, чтобы увеличить прибыль? Сократить затраты или увеличить долю рынка?»), будущее бизнес-аналитики за системами, которые могут давать ответы и рекомендации, а не большое количество знаний в виде отчетов. В результате на рынке появляется новая тенденция, называемая *адаптивной бизнес-аналитикой (adaptive business intelligence)*. Помимо выполнения роли традиционной бизнес-аналитики (преобразования данных в знания), адаптивная бизнес-аналитика также включает в себя процесс принятия решений, основанный на прогнозировании и оптимизации.

В то время как *бизнес-аналитика* часто определяется как «широкая категория прикладных программ и технологий для сбора, хранения, анализа и предоставления доступа к данным», *адаптивная бизнес-аналитика* — это дисциплина использования методов прогнозирования и оптимизации для создания самообучающихся систем принятия решений. Системы адаптивной бизнес-аналитики включают в себя элементы интеллектуального анализа данных (data mining), прогнозного моделирования, прогнозирования, оптимизации и адаптивности и используются бизнес-менеджерами для принятия более эффективных решений.

**Во второй главе** рассматривается создание *системы адаптивной бизнес-*

*аналитики*. Поскольку все организации работают в постоянно меняющейся среде, они вынуждены постоянно адаптироваться и приспосабливаться. Следовательно, для создания системы адаптивной бизнес-аналитики необходимо выполнить следующие четыре шага:

- Имеющиеся данные должны быть подготовлены и тщательно проанализированы (стадия интеллектуального анализа данных).
- Модель прогнозирования должна быть разработана на основе результатов интеллектуального анализа данных.
- Должен быть разработан модуль оптимизации, который использует модель прогнозирования, чтобы рекомендовать лучшее решение.
- Должен быть разработан модуль адаптивности, отвечающий за адаптацию модуля прогнозирования к изменяющейся во времени среде [2].

**В третьей главе** рассматривается понятие *науки о данных*. *Наука о данных (data science)* — междисциплинарная область, связанная с *большими данными (big data)* и *машинным обучением (machine learning)*, использующая научные процессы, методы и алгоритмы для извлечения идей и *бизнес-аналитики* из разнообразных неструктурированных и структурированных данных [3].

Рабочий процесс науки о данных включает в себя ряд сложных процессов, включая сбор данных, хранение данных, очистку данных, обработку данных, подготовку данных, кластеризацию данных, моделирование данных и обобщение аналитических данных.

**В четвертой главе** изучается *интеллектуальный анализ данных* и *статистика для принятия решений*. *Интеллектуальный анализ данных (data mining)* представляет собой слияние ряда других дисциплин, в первую очередь статистики и машинного обучения, применительно к проблеме извлечения информации из больших баз данных.

Современная дисциплина широко используется в бизнесе и других организациях. Описание особых характеристик коммерческих данных из ряда областей пользования служит иллюстрацией необычайной широты потенциальных применений [4].

Основными факторами успеха проекта реализации интеллектуального анализа данных являются:

- точные, существенные и реалистичные цели;
- богатство и, прежде всего, качество собранной информации;

- актуальность используемых методов интеллектуального анализа данных;
- удовлетворительный вывод сгенерированной информации и правильная интеграция в информационную систему, где это необходимо;
- анализ результатов и отзывы об опыте каждого приложения интеллектуального анализа данных.

**В пятой главе** рассматривается понятие машинного обучения, его использование, задачи, решаемые с помощью машинного обучения, и методы машинного обучения.

*Машинное обучение* заключается в извлечении знаний из данных. Это научная область, находящаяся на пересечении статистики, искусственного интеллекта и компьютерных наук и также известная как прогнозная аналитика или статистическое обучение. В последние годы применение методов машинного обучения в повседневной жизни стало обыденным явлением. Многие современные веб-сайты и устройства используют алгоритмы машинного обучения, начиная с автоматических рекомендаций по прослушиванию музыки, заказа еды или покупки продуктов, и заканчивая персонализированными онлайн-радиотрансляциями и распознаванием друзей на фотографиях [5].

Наиболее успешные алгоритмы машинного обучения — это те, которые автоматизируют процессы принятия решений путем обобщения известных примеров. В этих методах, известных как *обучение с учителем* или *контролируемое обучение* (*supervised learning*), пользователь предоставляет алгоритму пары объект-ответ, а алгоритм находит способ получения ответа по объекту. В частности, алгоритм способен выдать ответ для объекта, которого он никогда не видел раньше, без какой-либо помощи человека.

Алгоритмы машинного обучения, которые учатся на парах объект-ответ, называются алгоритмами обучения с учителем, так как «учитель» показывает алгоритму ответ в каждом наблюдении, по которому происходит обучение. Несмотря на то, что создание набора с объектами и ответами — это часто трудоемкий процесс, осуществляемый вручную, алгоритмы обучения с учителем интерпретируемы и качество их работы легко измерить. Если задачу можно сформулировать в виде задачи обучения с учителем, и можно создать набор данных, который включает в себя ответы, вероятно, машинное обучение решит эту проблему.

Также типом машинного обучения является *обучение с подкреплением*.

Задача обучения с подкреплением состоит в выработке системы (агента), которая улучшает свое качество на основе взаимодействий со *средой*. Поскольку информация о текущем состоянии среды, как правило, содержит и так называемый сигнал *вознаграждения*, обучение с подкреплением можно представить как область, имеющую отношение к обучению с учителем. Однако в обучении с подкреплением эта обратная связь является не меткой или значением, раз и навсегда определенными в результате прямых наблюдений, а мерой того, насколько хорошо действие было оценено функцией вознаграждения. В ходе взаимодействия со средой агент на основе разведочного подхода путем проб и ошибок или совещательного планирования может использовать обучение с подкреплением для вычленения серии действий, которые максимизируют это вознаграждение.

Еще один вид алгоритмов — это *алгоритмы обучения без учителя* или *алгоритмы неконтролируемого обучения* (*unsupervised algorithms*). В алгоритмах неконтролируемого обучения известны только объекты, а ответов нет [6].

**В шестой главе** рассматриваются нейронные сети и их обучение. Интеллектуальный анализ данных не был бы таким же без *нейронных сетей*, которые лежат в основе некоторых описательных и прогнозирующих методов интеллектуального анализа данных. Нейронные сети получили широкое распространение благодаря своей моделирующей способности (они могут аппроксимировать любую достаточно регулярную функцию) с отличными результатами в широком круге задач, даже при столкновении со сложными явлениями, неправильными формами и данными, которые трудно понять и обработать.

Алгоритмы машинного обучения нейронных сетей принято относить к области *глубокого обучения* (*deep learning*) — это построение и обучение нейронных сетей с большим числом слоев. Смысл алгоритма заключается в следующем: имеется набор данных и цель, которую нужно достигнуть, обучив алгоритм (например, предсказать число или отнести объект к определенному классу).

Простейшая нейронная сеть *перцептрон* (*perceptron*) представляет собой упрощенное отображение работы биологической сети, состоящей из нейронов и связей между ними. Эта модель стремится повторить восприятие информации человеческим мозгом и учитывает три этапа такого процесса:

- восприятие информации через сенсоры (входной слой);
- создание ассоциаций (скрытый слой);

— реакцию (выходной слой) [7].

Нейронная сеть с выбранными весами либо устанавливаемыми весами исходя из определенных математических соображений имеет место быть, когда сеть относительно небольшая. При увеличении числа нейронов и связей между ними ручной подбор становится невозможным, и возникает задача нахождения алгоритма, который бы автоматически подбирал эти связи. Такой процесс подбора связей и называют *обучением нейронной сети*.

Один из распространенных подходов к обучению заключается в последовательном предъявлении нейронной сети векторов наблюдений и последующей корректировке весовых коэффициентов так, чтобы выходное значение соответствовало как можно ближе требуемому отклику. Это и есть обучение с учителем, потому что для каждого вектора наблюдений известен желаемый результат. Соответственно, от сети требуется, чтобы она показывала этот желаемый результат.

Универсального алгоритма обучения нет, поэтому лучшее, что можно сделать — это выбрать тот алгоритм, который хорошо себя зарекомендовал в прошлом. Таковым является *алгоритм обратного распространения ошибки* (*back propagation algorithm*), который базируется на *алгоритме градиентного спуска* (*gradient descent*).

**В седьмой главе** представлен анализ программных средств для построения и обучения нейронных сетей.

Используя метод обратного распространения ошибки, алгоритм обучения и структуры сети описывается вручную. Но это не лучший путь, особенно, когда нейронная сеть становится большой, а критериев качества и методов оптимизации очень много. Гораздо проще воспользоваться одним из пакетов по разработке и обучению нейронных сетей.

*Библиотека машинного обучения Python* — это библиотека функций и методов, позволяющая легко масштабировать научные и числовые вычисления для оптимизации рабочих процессов машинного обучения. Фреймворки позволяют разработчикам создавать модели машинного обучения за меньшее время без необходимости вникать в тонкости базовых алгоритмов.

С библиотекой машинного обучения на основе Python разработчики и специалисты по обработке данных могут выполнять сложные задачи, не переписывая длинные фрагменты кода. Фреймворки предлагают более простой способ



определения, создания и развертывания моделей машинного обучения с помощью предварительно созданных компонентов библиотеки, которые упрощают машинное обучение и удобны для разработчиков [5].

Python — один из самых популярных и быстрорастущих языков программирования, превосходящий по производительности некоторые другие языки, такие как PHP, C#, JavaScript и Java. Согласно отчету компании Statista за февраль 2022 года, Python является третьим (48,24%) наиболее часто используемым языком программирования разработчиками по всему миру. Благодаря простоте и читабельности языка программисты предпочитают Python другим языкам.

**В восьмой главе** представлено описание программной реализации информационной системы для прогнозирования категориальных тенденций IT-рынка и прогнозирования вероятности успеха компаний на рынке. Был взят эксперимент с прогнозированием на 2015 год. Необходимо создать полносвязную нейронную сеть прямого распространения, используя алгоритм линейной регрессии. Линейная регрессия выполняет задачу по прогнозированию значения зависимой переменной на основе заданной независимой переменной.

Для проектирования нейронной сети был взят датасет из 914 наборов данных, содержащий в себе информацию о компаниях, появившихся в 1985-2014 годах. В нем представлены следующие характеристики данных: название компании, вложения в компанию на начало и конец года основания, вложения в компанию на начало и конец последнего года (конец последнего года существования компании, если она закрылась, или конец 2014 года, если компания продолжает существовать), количество партнеров компании, количество крупных вложений, количество этапов развития, попадание в известные рейтинги, среднее количество клиентов, категория компании, год основания компании, статус компании на 2015 год — существует ли она на рынке.

Для обработки данных использовалась библиотека Pandas, для обучения нейронной сети использовалась библиотека Keras, а для построения графиков — библиотека Matplotlib.

После считывания данных с датасета «dataset.csv» необходимо было разбить их на *тренировочную (training data)* и *тестовую (test data) выборки*. Обучающая выборка используется непосредственно для «обучения» той или иной модели, т. е. для построения математических отношений между некоторой переменной-откликом и предикторами, тогда как тестовая выборка служит

для получения оценки прогнозных свойств модели на новых данных, т. е. данных, которые не были использованы для обучения модели. Для тренировки были взяты компании, основанные в 1985-2011 годах, а для непосредственного прогнозирования — компании, основанные в 2012-2014 годах.

Далее была создана модель, в скрытый слой которой было добавлено два слоя нейронов. Первый слой состоит из 50 нейронов, второй состоит из 10, а выходной слой — из одного нейрона. У скрытого слоя функцией активации является *гиперболический тангенс*, у выходного слоя функция активации — *сигмоида*. В качестве параметров оптимизации обучения были взяты функция потерь *средняя квадратичная ошибка (MSE)*, оптимизатор *Adam*, метрика *средняя абсолютная ошибка (MAE)*.

Далее было запущено обучение нейронной сети. Были указаны входные и выходные данные, *эпоха (epoch)* — количество запусков алгоритма на всем наборе обучающих данных, равное 50, *пакет (batch)* — количество строк набора данных (сэмплов), которые необходимо взять для обновления параметров модели, равное 20. Для *валидационной выборки* было выбрано 30% от обучающей выборки.

Далее было запущено тестирование обученной модели на данных, включающих в себя компании, основанные в 2012-2014 годах, а затем на данных, включающих в себя компании, основанные до 2011 года включительно. Второе тестирование было запущено для дальнейшего анализа результатов на графиках.

Затем были созданы две пустые таблицы для каждого временного промежутка, куда были добавлены заранее запомненные наименования компаний, категории, годы основания, статусы существования компаний на 2015 год в столбец *status+* и добавлены новые предсказанные машиной результаты.

На рисунке 1 изображен график средней квадратичной ошибки (MSE) модели.

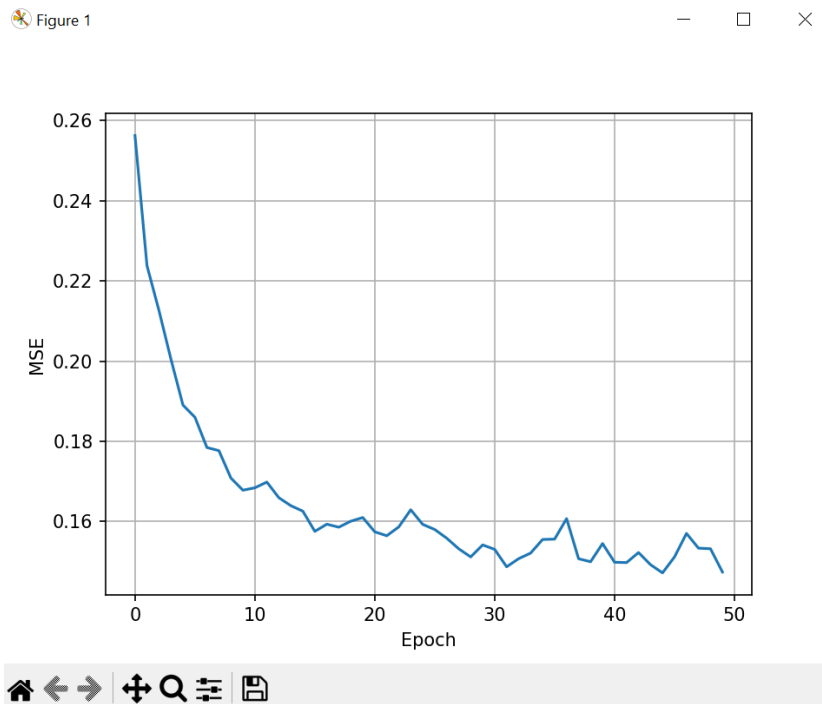


Рисунок 1 — График средней квадратичной ошибки (MSE)

На рисунке 2 изображен график средней абсолютной ошибки (MAE) модели.

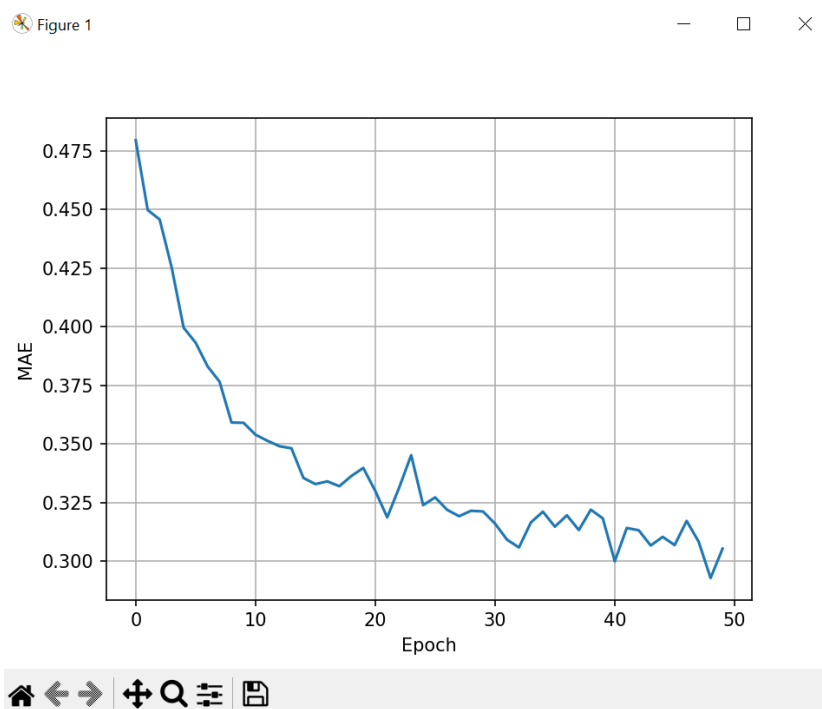


Рисунок 2 — График средней абсолютной ошибки (MAE)

Интерфейс приложения для пользователей на Python был разработан при помощи графической библиотеки Tkinter.

При запуске приложения пользователю предоставляется выбор функционала: поиск компаний по категории и/или году основания, загрузка таблицы с

результатом прогноза успеха для всех компаний, основанных в 2012-2014 годах, а также прогноз категорийных тенденций. Выбор функционала представлен на рисунке 3.

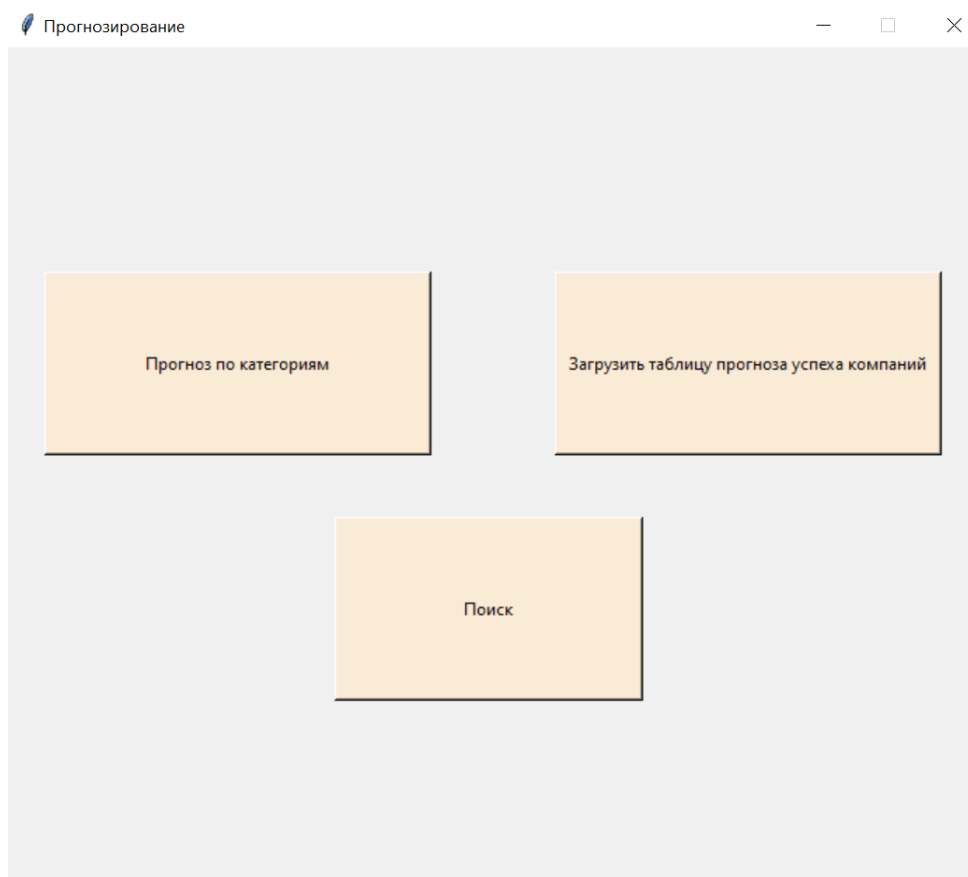


Рисунок 3 — Выбор функционала приложения

## ЗАКЛЮЧЕНИЕ

По мнению экспертов, заметной технологической тенденцией в ближайшие годы будет переход от фактографического анализа (в чем суть ситуации?) к предиктивному анализу (что может произойти?) и сценарному моделированию (что будет, если...?), а также дальнейшее совершенствование поиска зависимостей в данных и оценки их влияния на бизнес.

Внедрение информационных систем адаптивной бизнес-аналитики помогает минимизировать угрозы для развития компании, обеспечить прозрачность рабочих процессов и повысить их эффективность, помогая бизнесу выйти на новый уровень развития. Системы моделируют события будущего компании исходя из истории компании, текущей ситуации и ее динамики. На базе созданных моделей принимаются стратегические решения. Информационные системы предоставляют результаты анализа в доступном для восприятия виде и способствуют оперативному получению нужной информации для любого специалиста.

Аналитика с помощью BI-систем дает бизнесу базу для принятия решений с перспективой стабильного развития в будущем. Рекомендации BI-аналитиков помогают менеджменту получить четкие и актуальные сведения о текущей ситуации в компании и за ее пределами.

Компании стремятся включить в аналитическую обработку все больше данных разного типа, все большее количество сотрудников вовлекается в работу с данными, чтобы принимать еще более эффективные и своевременные решения. Данные должны оставаться актуальными, достоверными и единообразными, а также корректно использоваться всеми участниками процесса. Корректная работа сложной модели, особенно построенной с помощью машинного обучения, — проблема, которая пока не имеет полного решения.

Отраслевые специалисты видят широкие потенциальные возможности для использования в системах алгоритмов машинного обучения. Более того, интеллектуальные информационные технологии, обеспечивая человека умными вспомогательными инструментами, устанавливают новую планку качества работы аналитика. Опрос европейских компаний, проведенный в 2020 г., продемонстрировал высокий интерес корпоративных финансовых служб к планированию и прогнозированию с помощью машинного обучения. В свете этой тенденции спроса на автоматизацию прогнозной функции с применением методов машинного обучения усиливается.

В ходе работы была рассмотрена специальная литература по изучаемой проблеме, исследовано содержание ключевых понятий. Также были изучены методы и алгоритмы машинного обучения, проанализированы программные средства для построения и обучения нейронных сетей. Было реализовано прогнозирование с помощью нейронной сети и разработан графический интерфейс для пользователей. Цель бакалаврской работы достигнута, задачи выполнены в полном объеме.

**Основные источники информации:**

- 1 Michalewicz Z., Schmidt M., Michalewicz M., Chiriac C. Adaptive Business Intelligence / Z. Michalewicz, M. Schmidt, M. Michalewicz, C. Chiriac. — Springer, 2006. — 87 с. — Яз. англ.
- 2 Ханк Дж. Э., Райте А. Дж., Уичерн Д. У. Бизнес-прогнозирование / Дж. Э. Ханк, А. Дж. Райте, Д. У. Уичерн. — Издательский дом «Вильямс», 2003. — 55 с. — Яз. рус.
- 3 Грас Дж. Data Science. Наука о данных с нуля / Дж. Грас; пер. с англ. — СПб. : БХВ-Петербург, 2017. — 75 с. — Яз. рус.
- 4 Tuffery S. Data Mining and Statistics for Decision Making / S. Tuffery; translated by Rod Riesco. — John Wiley & Sons, 2011. — 83 с. — Яз. англ.
- 5 Рашка С. Python и машинное обучение / С. Рашка; пер. с англ. А. В. Логунова. — М. : ДМК Пресс, 2017. — 57 с. — Яз. рус.
- 6 Мюллер А., Гвидо С. Введение в машинное обучение с помощью Python / А. Мюллер, С. Гвидо. — Москва, 2016-2017. — 91 с. — Яз. рус.
- 7 Николенко С., Кадурын А., Архангельская Е. Глубокое обучение. Погружение в мир нейронных сетей / С. Николенко, А. Кадурын, Е. Архангельская. — СПб. : Питер, 2018. — 64 с. — Яз. рус.