

*Сидорин Александр Владимирович, магистр, Национальный исследовательский Мордовский государственный университет им. Н.П. Огарева*

*Юржонков Никита Владимирович, магистр, Национальный исследовательский Мордовский государственный университет им. Н.П. Огарева*

*Камаева Анастасия Андреевна, магистр, Национальный исследовательский Мордовский государственный университет им. Н.П. Огарева*

*Тепяева Юлия Евгеньевна, магистр, Национальный исследовательский Мордовский государственный университет им. Н.П. Огарева*

*Егунова Алла Ивановна, кандидат исторических наук, доцент кафедры автоматизированных систем обработки информации и управления, Национальный исследовательский Мордовский государственный университет им. Н.П. Огарева*

## **СИСТЕМА АНАЛИЗА И ПРЕДСКАЗАНИЯ ДИНАМИКИ ЦЕН**

**Аннотация:** В статье рассматриваются методы машинной обработки данных с помощью алгоритмов искусственного интеллекта. Приводится обзор существующих методов машинного обучения, а также приводится краткий обзор методов классификации. Рассматриваются преимущества и недостатки различных методов машинного обучения. Особое внимание в статье уделяется методам парсинга, классификации и предсказания на основе нейронных сетей. В статье описываются принципы построения системы прогнозирования и её основные этапы. Предложена процедура построения моделей для анализа и предсказания динамики цен. Проведен анализ эффективности применения полученных моделей для решения задач прогнозирования и принятия решений. Показано, что использование моделей, построенных на основе машинного

обучения, позволяет повысить эффективность прогнозирования в несколько раз по сравнению с использованием только методов статистического анализа. Рассматриваются методы выбора модели для построения нейронной сети, а также методы улучшения качества предсказания.

**Ключевые слова:** машинное обучение, прогнозирование, классификация, парсинг, ценообразование.

**Annotation:** The article discusses the methods of machine data processing using artificial intelligence algorithms. An overview of existing machine learning methods is provided, as well as a brief overview of classification methods. The advantages and disadvantages of various machine learning methods are considered. Particular attention is paid to the methods of parsing, classification and prediction based on neural networks. The article describes the principles of building a forecasting system and its main stages. A procedure for constructing models for analyzing and predicting price dynamics is proposed. An analysis of the effectiveness of the application of the obtained models for solving problems of forecasting and decision making is carried out. It is shown that the use of models built on the basis of machine learning makes it possible to increase the efficiency of forecasting by several times compared to using only statistical analysis methods. Methods for choosing a model for building a neural network, as well as methods for improving the quality of prediction, are considered.

**Keywords:** machine learning, prediction, classification, parsing, pricing.

## **Введение**

Прогнозирование является одним из важнейших направлений в современной науке. Во все времена предметная область прогнозирования и планирования остается одной из наиболее релевантных, так как предвидение близлежащей перспективы позволяет сократить риск и корректно управлять бизнес-процессами [1]. В основе процесса прогнозирования лежит построение модели, которая способна предсказать будущее поведение объекта.

Прогнозирование является важнейшим инструментом планирования, так как позволяет определить, каким образом может измениться ситуация в будущем, и спрогнозировать возможные последствия. На практике прогнозирование не всегда бывает точным, но, тем не менее, оно необходимо, так как помогает принять наиболее оптимальные управленческие решения [2].

### **1 Практическое значение разрабатываемой системы**

В интернете существует огромное множество сервисов с большим ассортиментом товаров, начиная от продажи продуктов и заканчивая криптовалютой и недвижимостью. В данный момент у пользователей появилась проблема с выбором нужного товара из большого множества. Пользователи вынуждены находить сразу несколько доверенных магазинов, запоминать цену товара, или же просто оставлять вкладку с ним, и искать сразу из нескольких вариантов самый выгодный. Но большинство пользователей не хотят заниматься таким однообразным процессом, поэтому просто совершают покупки в первом попавшемся на глаза магазине, или же просто в том, в котором совершают покупки постоянно, хотя вполне возможна ситуация, что где-то может быть более дешёвый вариант нужного товара.

Самые популярные сферы для анализа и предсказания цен:

- недвижимость;
- ценные бумаги;
- автомобили;
- акции;
- криптовалюты;

Формирование цены и её прогнозирование – это одна из приоритетных задач в современном обществе. В условиях рыночной экономики цена является важнейшим показателем формирования спроса и предложения на рынке.

В связи с этим возможное решение проблемы прогнозирования – разработка обучаемой системы анализа и предсказания изменения цен. Подобного рода система имеет широкое применение, в первую очередь в фирмах по продаже недвижимости или, например, в центрах налоговой

службы.

## **2 Методы и подходы к прогнозированию**

Для успешной разработки системы необходимо глубокое изучение предметной области, связанной в первую очередь с движениями активов. При этом необходимо учитывать, что в области финансов изучаются в первую очередь вопросы, связанные с экономикой, – экономические явления, процессы и отношения. Поэтому для разработки таких систем необходимо знание экономики и финансовых наук. Кроме того, для решения задач управления финансами особое значение имеет применение математических методов.

Большинство методов прогнозирования требует больших временных затрат и не дает точного прогноза. В связи с этим возникает необходимость в создании модели, которая могла бы наилучшим образом прогнозировать экономическое развитие.

Для ее решения необходимо определить, какие свойства или характеристики метода должны быть учтены при его выборе.

К таким свойствам относятся:

- точность прогноза;
- надежность прогноза;
- простота метода;
- экономичность метода;
- скорость получения результатов прогноза.

Так, например, в нашей стране в настоящее время широко используются различные модели, основанные на фундаментальном и техническом анализе, а также на различных статистических подходах к прогнозированию.

## **3 Техническое описание системы прогнозирования цен**

Основными этапами проектирования системы являются:

- 1) парсинг информации;
- 2) классификация, анализ данных(классификатор);
- 3) прогнозирование (нормализация данных).

1. Парсер – это программа, осуществляющая парсинг, предназначенная

для автоматического анализа и извлечения данных из определенного формата данных, чтобы в последствии предоставить эти данные пользователю.

Рассмотрим методику сбора данных с сайта. Существует несколько способов для реализации извлечения данных с web-сайтов. Далее рассмотрим несколько методик, наиболее применяющихся для данной задачи.

Суть первого метода состоит в извлечении данных, используя только код HTML-документа. Данный подход реализуется следующим образом: берется исходный код web-страницы, извлекаются необходимые данные, происходит выделение текста из гипертекстовой разметки путем использования регулярных выражений и специальных библиотек. Затем происходит сохранение полученных данных.

Второй метод заключается в реализации сбора данных при помощи компьютерного зрения. Производится рендеринг страницы, затем происходит ее преобразование в изображение и распознавание контента, а также его дислокации.

Третий метод заключается в том, чтобы полностью извлечь информацию на уровне web-сайтов. При помощи специальных формул, схем, таблиц и приложений возможно детектировать на страницах нужный фрагмент кода с необходимыми данными. Данный метод позволяет сделать это автоматически, что позволит сэкономить время и предупредить возможные ошибки копирования данных.

2. Классификация данных – это процесс организации данных по категориям, которые облегчают их извлечение, сортировку и хранение для дальнейшего использования. На базовом уровне процесс классификации облегчает поиск и извлечение данных.

Классификация данных является важной частью управления жизненным циклом данных, которая определяет, к какой стандартной категории или группировке относится объект данных. Классификация данных включает в себя пометку данных, чтобы сделать их легко доступными для поиска и отслеживания. Классификация данных используется для категоризации

структурированных данных, что особенно важно для получения максимальной отдачи от неструктурированных данных. Категоризация данных также помогает идентифицировать повторяющиеся копии данных. Устранение избыточных данных позволяет снизить затраты на хранение и резервное копирование, что способствует эффективному использованию хранилища и максимизирует меры безопасности данных.

3. Нормализация – это масштабирование данных из диапазона так, чтобы все значения находились в диапазоне от 0 до 1. Процесс нормализации данных крайне важен в процессе машинного обучения, так как позволяет представить информацию в единообразном виде.

Рассмотрим алгоритм нормализации данных, используя класс `MinMaxScaler` модуля Python для машинного обучения `Scikit-learn`. Передовая практика использования `MinMaxScaler` и других методов масштабирования выглядит следующим образом:

1. необходимо подобрать значения масштабирования с помощью доступных обучающих данных. Так обучающие данные будут использоваться для оценки минимального и максимального наблюдаемых значений посредством вызова функции `fit()`;

2. необходимо применить значения масштабирования к обучающим данным. Так можно использовать нормализованные данные для обучения вашей модели посредством вызова функции `transform()`;

3. необходимо применить значения масштабирования к прогнозируемым данным. Так можно подготовить новые данные в будущем, на основе которых необходимо делать прогнозы.

Для реализации проекта по созданию системы прогнозирования необходимо определить источник данных, который будет использоваться в качестве основы для прогнозирования. На первом этапе происходит захват, регистрация и обработка данных. Данные захватываются посредством парсинга различных источников. После этого следует проведение анализа полученной информации (в том числе ее систематизация). Классификация осуществляется

по различным признакам. Проанализировав всю информацию, необходимо выбрать наиболее релевантную, а затем осуществить прогноз. Для этого проводится прогнозирование по нескольким параметрам [3].

#### 4 Теоретические аспекты нейронных сетей

Искусственный интеллект является одним из наиболее значимых направлений современной науки, которое позволяет решать задачи, стоящие перед человечеством. Его развитие происходит стремительными темпами. На данный момент все большую популярность при построении систем для прогнозирования приобретают нейросетевые технологии.

Под нейронными сетями понимается математическое отображение биологической модели сети взаимосвязанных нейронов. Нейросетевая структура получает на вход некоторый набор данных и в процессе корректировки своих параметров в результате формирует совокупность выходных сигналов.

Изображение, представленное ниже на рисунке 1, отражает архитектуру сетей. Модель представляет собой три основных блока. Уровень подающейся на вход информации, скрытый уровень нейронов и уровень выходных данных. Подобную инфраструктуру называют упреждающей сетью.

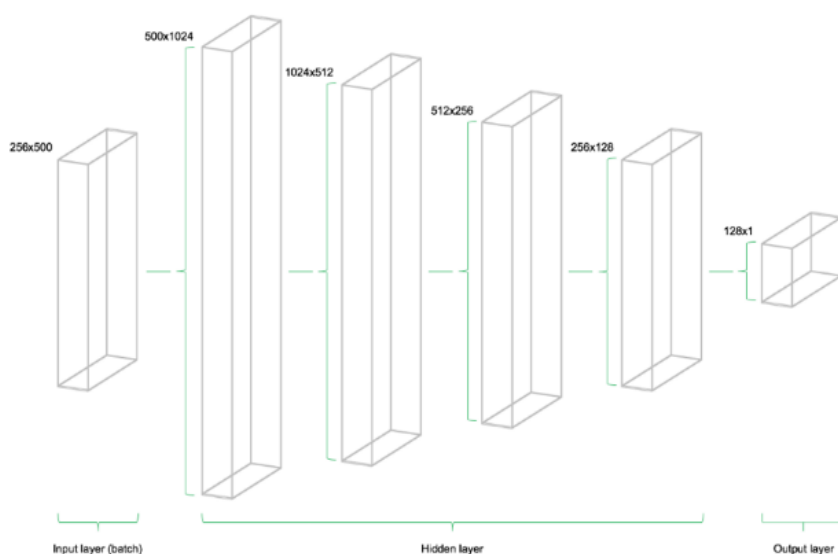


Рисунок 1 – Архитектура нейронной сети

Рассмотрим библиотеки машинного обучения для языка программирования Python.

1. Scikit-learn. Scikit-learn – один из наиболее широко используемых пакетов Python для Data Science и Machine Learning. Проект был начат в 2007 году Дэвидом Курнапо как проект Google Summer of Code, и с тех пор любой доброволец мог внести свой вклад. В настоящее время он обслуживается командой добровольцев.

2. Keras. Keras – это библиотека для языка программирования Python, которая предназначена для глубокого машинного обучения.

Преимущества Keras:

– Простота. Keras снижает когнитивную нагрузку на разработчиков, позволяя сосредоточиться на действительно важных частях проблемы.

– Гибкость. Keras использует принцип постепенного раскрытия сложности: простые рабочие процессы должны быть быстрыми и легкими, в то время как произвольно расширенные рабочие процессы должны быть реализованы через четкий путь, основанный на том, что уже изучено.

– Мощный. Keras обеспечивает лучшую в отрасли производительность и масштабируемость: он используется организациями и компаниями, включая NASA, YouTube или Waymo.

### **Заключение**

Исследование построения системы поддержки принятия решений на основе методов машинного обучения позволяет помочь провести технический анализ и прогнозирования изменения цен на рынке [4].

Использование навыков математического анализа и собственного опыта позволит пользователю строить более точные прогнозы и выстроить оптимальную стратегию. Нейросетевую модель можно настраивать и корректировать, изменяя выборку исторических данных, размер обучающей и тестируемой выборок чтобы добиться оптимальной точности прогноза. При решении этой задачи мы также выясняем, какие характеристики наиболее весомые в определении и прогнозировании цены.



### **Библиографический список:**

1. Афонин В. В. Основы анализа систем массового обслуживания: Учеб. пособие: Для студентов вузов, обучающихся по направлению 654600 "Информатика и вычисл. техника", специальности 220200 "Автоматизир. системы обраб. информ. и упр." / В. В. Афонин, С. М. Мурюмин, С. А. Федосин, В. В. Афонин, С. М. Мурюмин, С. А. Федосин. – Саранск: Изд-во Морд. ун-та, 2003. – 234 с. – ISBN 5-7103-0930-3.

2. Афонин В. В. Моделирование систем / В. В. Афонин, С. А. Федосин; Интернет-Университет Информационных Технологий. – Москва: ООО "Издательство "БИНОМ. Лаборатория знаний", 2010. – 231 с. – ISBN 978-5-9963-0352-6.

3. Афонин В. В. Анализ и моделирование типовых систем массового обслуживания: Учебное пособие для студентов по направлению подготовки «Инфокоммуникационные технологии и системы связи» / В. В. Афонин, В. В. Никулин. – Москва; Вологда: Общество с ограниченной ответственностью "Издательство "Инфра-Инженерия", 2023. – 232 с. – ISBN 978-5-9729-1187-5.

4. Ерофеев В. Т. Методы обработки экспериментальных данных / В. Т. Ерофеев, И. Н. Максимова, В. В. Афонин; Национальный исследовательский Мордовский государственный университет им. Н. П. Огарева, Пензенский государственный университет архитектуры и строительства. – Саранск: Национальный исследовательский Мордовский государственный университет им. Н.П. Огарёва, 2019. – 140 с. – ISBN 978-5-7103-3735-6.