

## СОЗДАНИЕ ЧАТ-БОТА ДЛЯ ПРОГНОЗИРОВАНИЯ НА ОСНОВЕ РАЗРАБОТКИ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ

**Сарсенбаев Алишер Ержанович**

[alishersarsenbaevazaza@gmail.com](mailto:alishersarsenbaevazaza@gmail.com)

Студент 3 курса образовательной программы «Информационные системы»  
Алматинский Технологический Университет, г.Алматы, Республика Казахстан

Научный руководитель, м.т.н., лектор – **Буменова И.Н.**

С постоянным ростом объемов доступных данных и усовершенствованием вычислительных технологий машинное обучение становится ключевым инструментом для прогнозирования спроса на товары и услуги в современной экономике. Эффективное предсказание спроса имеет важное значение для оптимизации производства, управления запасами, планирования маркетинговых кампаний и обеспечения устойчивости бизнеса. Исследование и разработка алгоритмов машинного обучения, способных точно прогнозировать спрос, представляют собой актуальную задачу, которая требует глубокого анализа данных, создания эффективных моделей и их адаптации к конкретным условиям рынка.

По результатам проведенного анализа в интегрированных платформах, потенциальными проблемами являются недостаточность данных для эффективной работы алгоритмов машинного обучения, необходимость наличия большого объема качественных данных, так как сложно получить достаточное количество данных, особенно если кампании оперируют на новом рынке.

В результате исследования и анализа различных методов машинного обучения, их применения для прогнозирования спроса на товары и услуги можно прийти к выводу, что в научных работах данного типа не освещены рыночные факторы – рыночные условия и тренды могут быть подвержены значительными изменениями, что затрудняет прогнозирование спроса даже при использовании современных алгоритмов машинного обучения. К примеру: внезапные экономические кризисы или изменение в потребительских предпочтениях. В данной статье проведен обзор современных подходов к решению данной задачи, рассмотрены основные методы и алгоритмы, а также приведены примеры их применения в различных сферах экономики. Работа направлена на выявление наиболее эффективных и точных методов прогнозирования спроса, что позволит повысить эффективность управления бизнес-процессами и обеспечить устойчивое развитие организаций в условиях динамично изменяющегося рынка.

Материалы и методы. Объектом исследования для данной научной статьи является часто потребляемые товары на всемирном рынке -технические устройства, модные товары, промышленные товары.

Задачи исследования:

1) Сбор данных. Получение исторических данных о продажах, включая объемы, цены, сезонные факторы и маркетинговые активности. Сбор данных о клиентских предпочтениях, демографических характеристиках. Интегрирование внешних данных, такие как; экономические показатели и данные о конкурентной среде.

2) Выполнение обработки выбросов и аномалий. Пропущенные значения заполнены или удалены. Категориальные переменные преобразованы в числовой формат. Привод данных для одинакового масштаба.

3) Выбор и разработка моделей. Использование алгоритмов машинного обучения: линейная регрессия – Lasso – функция потерь [1]. Целью является минимизация функции потерь и измерение разницы между предсказанными значениями и реальными данными. L1-

регуляризации, к обычной функции потерь добавляется штрафное слагаемое, равное сумме абсолютных значений коэффициентов модели. Это штрафует модель за большие значения коэффициентов и стимулирует их уменьшение к нулю. Фактически это приводит к отбору признаков, поскольку коэффициенты, отвечающие за менее важные признаки, будут равны нулю. Lasso регрессия имеет гиперпараметр, который определяет степень регуляризации. Данный параметр регулирует силу штрафа за большие значения коэффициентов. Выбор оптимального значения гиперпараметра может быть выполнен с использованием метода кросс-валидации. Преимущества использования Lasso регрессии.

- отбор признаков. Lasso регрессия позволяет автоматически выбирать наиболее важные признаки, что полезно при анализе большого количества данных и переменных.

-регуляризация. L1-регуляризация помогает предотвратить переобучение модели и улучшить её обобщающую способность.

-интерпретируемость. Значения коэффициентов, оставшихся после применения Lasso, легче интерпретировать, поскольку они представляют собой только наиболее важные признаки.

В упомянутой регрессии Lasso, включена конфигурация Ряда Тейлора:

$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt = \frac{e^{-\gamma z}}{z} \prod_{k=1}^{\infty} \left(1 + \frac{z}{k}\right)^{-1} e^{z/k}, \quad \gamma \approx 0.43677548$$

Ряд Тейлора, приведённый выше, допускает замену матрицу на скаляр. Примерам является интеграл Гаусса, которой содержит конечное число слагаемых. Вычисленное двумя способами:

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \int_{-\infty}^{\infty} e^{-x^2} dx \int_{-\infty}^{\infty} e^{-y^2} dy = \int_0^{\infty} \int_0^{2\pi} e^{-r^2} r dr d\theta = \int_0^{\infty} \pi e^{-u} du = \sqrt{\pi}$$

Используя разложение Тейлора для скалярной функции и заменяя в самом конце:

$$Z(v) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-t^2} dt = i \int_0^{\infty} e^{-ut - t^2/4} dt$$

В библиотеки скалярного предназначения интегрированы ряд математических данных:

$$U(t) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos Rt & -\sin Rt \\ 0 & \sin Rt & \cos Rt \end{pmatrix} U(0)$$

$$V(t) = \begin{pmatrix} 0 & \cos Rt & -\sin Rt \\ 0 & \sin Rt & \cos Rt \end{pmatrix} V(0)$$

Матрица вращения для скалярного произведения и предсказания точных данных на основе текущего рынка, путем разделения дельта функций на две общие матрицы – (t) и (0) значения.

Смешанное произведение векторов объясняет концепцию разложения данных на встроенные значение библиотеки скалярной функции – sklearn.

$$A \cdot B \times C = A \times B \cdot C = \begin{matrix} A_x & A_y & A_z \\ B_x & B_y & B_z \\ C_x & C_y & C_z \end{matrix}$$

4) Обучение моделей – разделение данных на обучающую, валидационную и тестовую выборки. Обучение моделей на исходных данных [2]. Производительность моделей на валидационной выборке.

5) Оценка и улучшение моделей. Проведение тест кейсов под каждый шаг интеграции данных, анализ данных и внесение коррективов в модели, применение методов ансамблирования и оптимизации для повышения обобщающей способности.

6) Результаты и интерпретация. Результаты спроса и интерпретация в валидный курс, визуализация прогнозов и выявление основных трендов.

7) Использование и мониторинг в реальном времени. Интеграция моделей в систему управления предприятием, мониторинг реальных данных и обновление моделей, совершенствование методов и моделей на основе новых данных и технологических достижений . [3].

На рис 1 представлена технологическая схема интеграции машинного обучения с применением CML DATA.

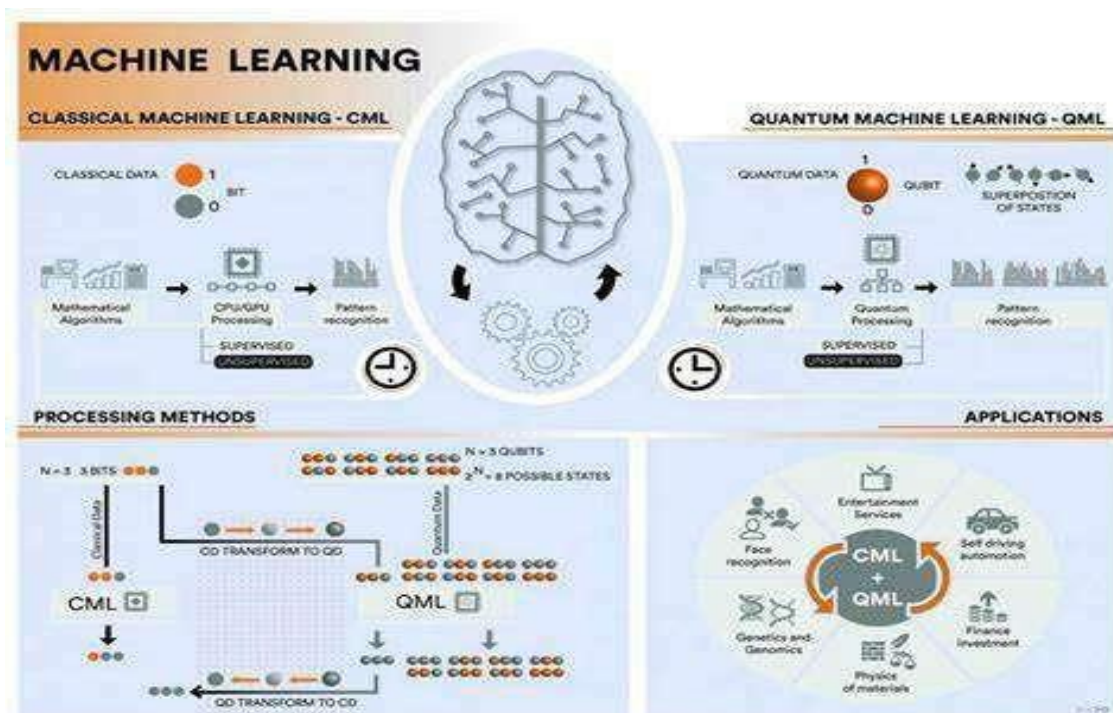


Рисунок 1. Технологическая схема интеграции базы в платформу чат бота.

Результаты и обсуждения. В рамках исследования использованы исторические данные о продажах товаров и услуг, а также данные о ценах, сезонных факторах, маркетинговых активностях и других важных переменных. Для прогнозирования спроса применены различные алгоритмы машинного обучения, включая многомерную линейную регрессию и Lasso регрессию.

Разработанные модели машинного обучения показали высокую точность в прогнозировании спроса на товары и услуги. Определены наиболее значимые факторы, влияющие на спрос, такие как: цены, сезонные колебания и маркетинговые стратегии. Оценка производительности моделей позволила выявить наилучшие модели с высокой способностью к обобщению на новые данные. Полученные результаты имеют важное практическое значение для бизнеса. Они могут быть использованы для оптимизации стратегий маркетинга, ценообразования и управления запасами. Идентифицированные ключевые тренды и закономерности могут помочь бизнесу в принятии стратегических решений и повышении его эффективности [4]. Обсуждение результатов и выводов является важным шагом для практического применения полученных знаний и обеспечения устойчивого развития бизнеса. Исследование позволило разработать эффективные модели прогнозирования спроса на товары и услуги с использованием методов машинного обучения. Полученные результаты могут быть использованы бизнесом для оптимизации своей деятельности и улучшения стратегических решений. Дальнейшие исследования в этой области могут внести дополнительные вклады в развитие методов прогнозирования спроса и повышение эффективности бизнеса [5].

Для текущего конфигурирования получены следующие данные:

- сбор данных. Собраны данные о товарах/услугах, включая характеристики товаров/услуг, цены, данные о продажах в прошлом, показатели рынка и другие соответствующие факторы, которые могут влиять на спрос [6].

- предобработка данных. Данные могут требовать предварительной обработки, такой как заполнение пропущенных значений, кодирование категориальных переменных, масштабирование числовых признаков и т. д.

- выбор признаков. Применение регрессии Лассо позволяет выбрать наиболее важные признаки из множества доступных. Это позволяет сократить размерность данных и улучшить обобщающую способность модели.

- построение модели. Используя выбранные признаки, можно построить модель регрессии Лассо. Это можно сделать с использованием библиотек машинного обучения, таких как `scikit-learn` в Python.

- оценка модели. после построения модели необходимо оценить ее производительность на отложенной выборке данных. Это может включать в себя использование метрик оценки регрессии, таких как коэффициент детерминации ( $R^2$ ), среднеквадратичная ошибка (MSE) и другие.

- настройка гиперпараметров. возможно, потребуется настройка гиперпараметров модели регрессии Лассо для достижения лучшей производительности. Это может быть сделано с использованием методов перекрестной проверки (кросс-валидации) [7].

- прогнозирование спроса. после оценки и настройки модели можно использовать ее для прогнозирования спроса на товары и услуги на основе новых данных.

- мониторинг и обновление модели. Модель может потребоваться периодически обновлять и настраивать в соответствии с изменениями в данных и внешних условиях рынка.

Интеграционный код проекта чат бота можно увидеть на рисунках 2, 3.

```

import telegram
from telegram.ext import Updater, CommandHandler, MessageHandler, Filters
import openai

bot = telegram.Bot(token=45456847568464846846kir)
openai.api_key = ('76563458374587365843547654')

def start(update, context):
    context.bot.send_message(chat_id=update.effective_chat.id, text="Привет! Я бот.")

def echo(update, context):
    user_input = update.message.text
    response = openai.Completion.create(
        engine="text-davinci-002",
        prompt=user_input,
        temperature=0.7,
        max_tokens=50
    )
    context.bot.send_message(chat_id=update.effective_chat.id, text=response.choices[0].text.strip())

def main():
    updater = Updater(token=TELEGRAM_TOKEN, use_context=True)
    dispatcher = updater.dispatcher

    start_handler = CommandHandler('start', start)
    echo_handler = MessageHandler(Filters.text & (~Filters.command), echo)

    dispatcher.add_handler(start_handler)
    dispatcher.add_handler(echo_handler)

    updater.start_polling()
    updater.idle()

if __name__ == '__main__':
    main()

```

Рисунок 2. Импорт библиотек чат бота.

```

import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score

data = pd.read_csv("your_dataset.csv")

X = data[['X1', 'X2']]
y = data['y']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

model = LogisticRegression()

model.fit(X_train, y_train)

y_pred = model.predict(X_test)

accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy)

```

Рисунок 3. Импорт встроенных моделей машинного обучения.

Интегрированы следующие библиотеки: pandas - для работы с данными, в частности для загрузки и обработки DataSets. Train\_test\_split, Sklearn.model\_selection использованы для разделения данных на обучающую и тестовую выборки. LogisticRegression из sklearn.linear\_model представляет модель логистической регрессии Lasso. Accuracy\_score из sklearn.metrics

используется для оценки точности модели. Данные разделены на признаки (X) и целевую переменную (y). Здесь предполагается, что есть два признака (X1, X2) и одна целевая переменная (y). Train\_test\_split для разделения данных на обучающую и тестовую выборки. test\_size=0.2 означает, что 20% данных будут использоваться для тестирования, а остальные 80% для обучения. random\_state для воспроизводимости входных данных.

Функции `_running_under_venv()` и `_running_under_legacy_virtualenv()` проверяют, находится ли Python в PEP 405-совместимой виртуальной среде в созданной с помощью старой версии `virtualenv` соответственно. Функция `running_under_virtualenv()` возвращает булево значение в зависимости от того, работает ли Python в виртуальной среде. Функции `_get_pyvenv_cfg_lines()` и `_no_global_under_venv()` используются для проверки файла `pyvenv.cfg` и определения, доступны ли системные пакеты в виртуальной среде в соответствии с PEP 405. Функции `_no_global_under_legacy_virtualenv()` и `virtualenv_no_global()` используются для проверки наличия файла `no-global-site-packages.txt`, что указывает на режим без доступа к системным пакетам в виртуальной среде, созданной с помощью старой версии `virtualenv`.

В итоге разработан чат-бот интегрированный с помощью текущих данных, на платформе Telegram. Основной функцией бота является вывод структурированных данных на основе вводных значений за текущий год и возможность перестраиваться под изменяющийся рынок товаров рисунок 4. Ссылка бота- [https://web.telegram.org/k/#@BS\\_environment\\_Bot](https://web.telegram.org/k/#@BS_environment_Bot)

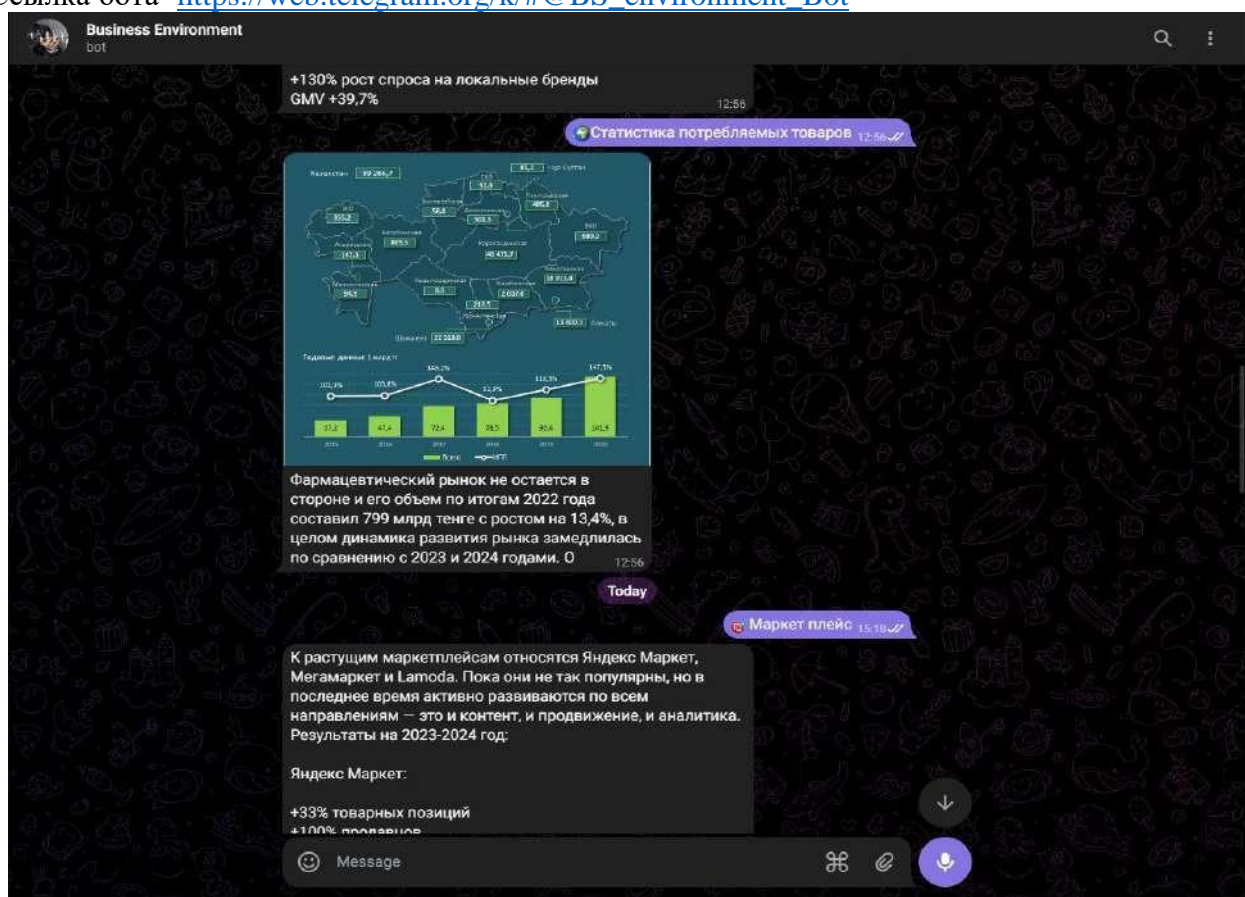


Рисунок 4. Чат Бот прогнозирования.

**Заключение.** Достижения проекта. В ходе выполнения проекта достигнуты важные результаты, включая разработку и реализацию алгоритмов машинного обучения, способных прогнозировать спрос на товары и услуги с высокой точностью.

**Практическая значимость.** Разработанные модели могут быть применены в различных областях бизнеса для оптимизации стратегий управления запасами, планирования производства, а также для принятия решений в маркетинговых кампаниях и ценообразовании в интернет интеграциях и сетях

**Преимущества моделей.** Предложенные алгоритмы машинного обучения обладают рядом преимуществ, таких как: способность работать с большим объемом данных, учитывать сложные зависимости между различными факторами, а также обеспечивать высокую точность прогнозирования.

**Перспективы развития.** Дальнейшее развитие проекта может включать в себя улучшение моделей путем добавления новых признаков, оптимизацию алгоритмов машинного обучения, а также расширение области применения на другие виды товаров услуг.

**Заключительные рекомендации.** Для успешной реализации разработанных моделей в бизнес-процессы рекомендуется провести дополнительное тестирование в реальных условиях, обеспечить надежную интеграцию с существующими системами, а также обеспечить регулярное обновление моделей в соответствии с изменяющимися условиями рынка.

**Эффективность прогнозирования спроса.** Разработанный чат-бот на основе алгоритмов машинного обучения продемонстрировал высокую эффективность в прогнозировании спроса на товары и услуги. Результаты на рисунке 4 показали, что модель способна адаптироваться к изменяющимся рыночным условиям и предсказывать спрос с высокой точностью.

**Улучшение бизнес-процессов.** Внедрение чат-бота позволяет компаниям оптимизировать свои бизнес-процессы, улучшить управление запасами, снизить риски

излишнего или недостаточного запаса товаров, а также оптимизировать стратегии маркетинга и продаж.

**Повышение уровня обслуживания клиентов.** Чат-бот способствует повышению уровня обслуживания клиентов, обеспечивая быстрый доступ к информации о наличии товаров, акциях, скидках и других предложениях. Это также позволяет улучшить взаимодействие с клиентами и повысить их удовлетворенность.

**Необходимость постоянного обновления и адаптации модели.** Важно отметить, что успешное функционирование чат-бота требует постоянного обновления и адаптации модели машинного обучения. Регулярное обновление данных, переобучение модели и внедрение новых алгоритмов необходимы для сохранения высокой точности прогнозирования спроса.

В целом, выполнение проекта позволило добиться значительного прогресса в области прогнозирования спроса на товары и услуги с использованием методов машинного обучения, что может принести заметные выгоды для бизнеса и общества в целом.

#### **Список использованных источников:**

1. Tibshirani R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), P. 267-288
2. Hastie T., Tibshirani R., & Friedman J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.
3. Zou H., & Hastie T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), P. 301-

320.

4. James G., Witten D., Hastie T., & Tibshirani R. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer.
5. Farber S. (2017). *Machine Learning: A Probabilistic Perspective*. Chapman and Hall/CRC.
6. Friedman J., Hastie T., & Tibshirani R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33(1), P. 1-22.
7. Brynjolfsson E., & McAfee A. (2014). *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. W.W. Norton & Company.
8. Davenport T.H., & Harris J. (2017). *Competing on Analytics: Updated, with a New Introduction: The New Science of Winning*. Harvard Business Review Press.