

УДК 004.8

## **СОЗДАНИЕ МОДЕЛИ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ ПОГОДЫ**

***Киселёв А.А.***

*Бакалавр,*

*Санкт-Петербургский Государственный Университет Промышленных  
Технологий и Дизайна, Высшая Школа Технологий и Энергетики,*

*Санкт-Петербург, Россия*

***Габдуллин Э.Х.***

*Бакалавр,*

*Санкт-Петербургский Государственный Университет Промышленных  
Технологий и Дизайна, Высшая Школа Технологий и Энергетики,*

*Санкт-Петербург, Россия*

***Москаленко П.А.***

*Бакалавр,*

*Санкт-Петербургский Государственный Университет Промышленных  
Технологий и Дизайна, Высшая Школа Технологий и Энергетики,*

*Санкт-Петербург, Россия*

***Леонова Н.Л.***

*Старший преподаватель кафедры прикладной математики и информатики,*

*Санкт-Петербургский Государственный Университет Промышленных  
Технологий и Дизайна, Высшая Школа Технологий и Энергетики,*

*Санкт-Петербург, Россия*

### **Аннотация**

В данной работе содержится информация о разработке искусственного интеллекта при помощи машинного обучения; о получении, обработке,

форматировании и сортировке начальных данных; о визуализации используемых данных; о практическом применении искусственного интеллекта в жизни человека. В научной работе рассмотрены примеры применения машинного обучения в метеорологии, а результатом работы является искусственный интеллект, который способен предсказывать погодные условия, создавая прогнозы из полученных данных.

**Ключевые слова:** искусственный интеллект, машинное обучение, погода, прогнозирование, график.

### ***CREATING A MACHINE LEARNING MODEL FOR WEATHER FORECASTING***

***Kiselev A.A.***

*Bachelor,*

*Saint Petersburg State University of Industrial Technologies and Design, Higher School of Technology and Energy,*

*Saint Petersburg, Russia*

***Gabdullin E.H.***

*Bachelor,*

*Saint Petersburg State University of Industrial Technologies and Design, Higher School of Technology and Energy,*

*Saint Petersburg, Russia*

***Moskalenko P.A.***

*Bachelor,*

*Saint Petersburg State University of Industrial Technologies and Design, Higher School of Technology and Energy,*

*Saint Petersburg, Russia*

***Leonova N.L.***

*Senior Lecturer of the Department of Applied Mathematics and Computer Science,*

*Saint Petersburg State University of Industrial Technologies and Design, Higher School of Technology and Energy,*

*Saint Petersburg, Russia*

## Abstract

This paper contains information about the development of artificial intelligence using machine learning; about the receipt, processing, formatting and sorting of initial data; about the visualization of the data used; about the practical application of artificial intelligence in human life. In the scientific work, examples of the use of machine learning in meteorology are considered, and the result of the work is an artificial intelligence that is able to predict weather conditions, creating forecasts from the data obtained.

**Keywords:** artificial intelligence, machine learning, weather, forecasting, graph.

В современных реалиях всё большей популярностью пользуется машинное обучение, а внедрение искусственного интеллекта в различные сферы деятельности человека. Искусственный интеллект (ИИ) становится сложнее, с каждым днём он выполняет всё более трудные задачи. Одной из областей применения ИИ - является прогнозирование. В данный момент прогнозирование становится всё более и более востребованной задачей современного общества.

Прогнозирование с помощью ИИ применяется в различных сферах деятельности, оно стало неотъемлемой частью жизни современного общества. Например, в медицине ИИ прогнозирует вероятность сердечного приступа у пациента, в экономике ИИ способен спрогнозировать рост или падение акций на фондовом рынке, ИИ может предсказывать погоду в ближайшие часы с высочайшей точностью.

В ходе исследования возможностей применения ИИ возникла теоретическая задача предсказания погодных условий в некоторой заданной местности. Таким образом, задача сводится к созданию самообучаемой программы, способной с высокой точностью определить погодные условия.

Подобные программы не являются новаторскими, ИИ уже успешно используется в течении нескольких лет в метеорологии. К примеру компания DeepMind на данный момент разрабатывает инструмент машинного обучения, который может вывести этот процесс на новый уровень точности. Специалисты используют высокоточные радиолокационные данные, которые отслеживают осадки каждые пять минут с разрешением один километр. Исследователи применили подход, известный как генеративное моделирование. Такая модель анализирует последние 20 минут наблюдений радара и создаёт прогноз на ближайшие 90 минут вперёд. Этот инструмент используется для прогнозирования средних и сильных дождей, оказывающих наибольшее воздействие на жизнь людей и экономику. Разработчики выяснили, что их продукт выполняет свою работу лучше, чем модели, используемые сегодня. Более 40 метеорологов из Метеорологического бюро Великобритании оценивали работу нового искусственного интеллекта. В 89% случаев "Глубинная генеративная модель дождя" от DeepMind была оценена выше, чем широко используемые сегодня методы прогнозирования текущей погоды.

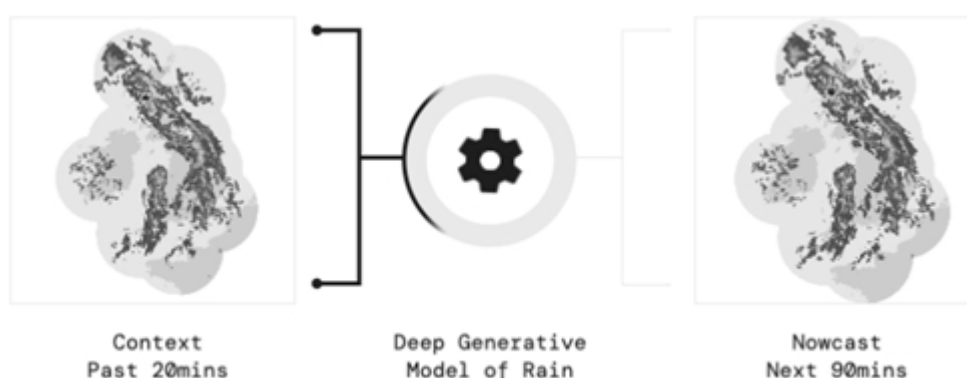


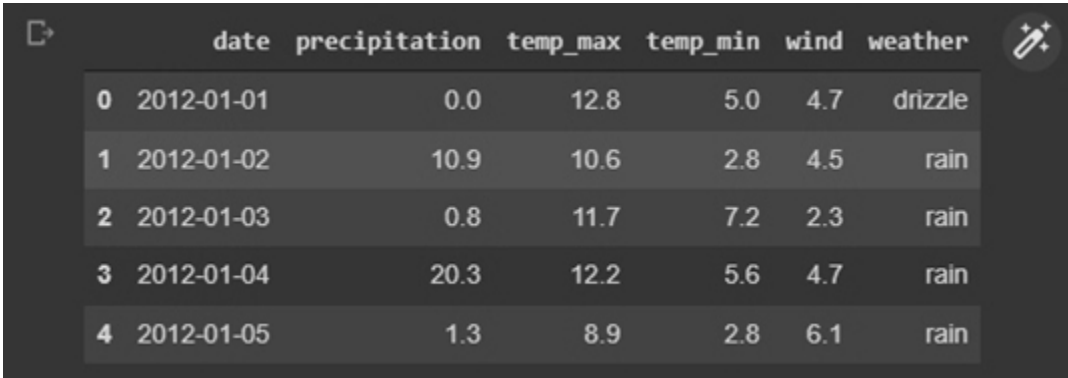
Рис.1. ИИ-метеоролог [16]

Работа над ИИ-метеорологом в компании DeepMind всё ещё продолжается. Разработчики увеличивают процент точности его прогнозов, а также обучают

создавать долгосрочные прогнозы погоды и предсказывать редкие и сильные дожди.

В нашем случае можно использовать более простые в разработке, но менее ресурсозатратные методы.

Изначально из открытых источников в интернете была получена база данных с такими признаками, как дата, выпадение осадков, максимальная температура, минимальная температура, ветер, тип погоды. На данном этапе были описаны библиотеки, которые впоследствии будут использованы в программе. Следующим шагом в программу была загружена база данных.



The image shows a screenshot of a database table with the following columns: date, precipitation, temp\_max, temp\_min, wind, and weather. The data is as follows:

	date	precipitation	temp_max	temp_min	wind	weather
0	2012-01-01	0.0	12.8	5.0	4.7	drizzle
1	2012-01-02	10.9	10.6	2.8	4.5	rain
2	2012-01-03	0.8	11.7	7.2	2.3	rain
3	2012-01-04	20.3	12.2	5.6	4.7	rain
4	2012-01-05	1.3	8.9	2.8	6.1	rain

Рис. 2. Часть базы данных [Авторская разработка]

Далее программе было необходимо изучить данные.

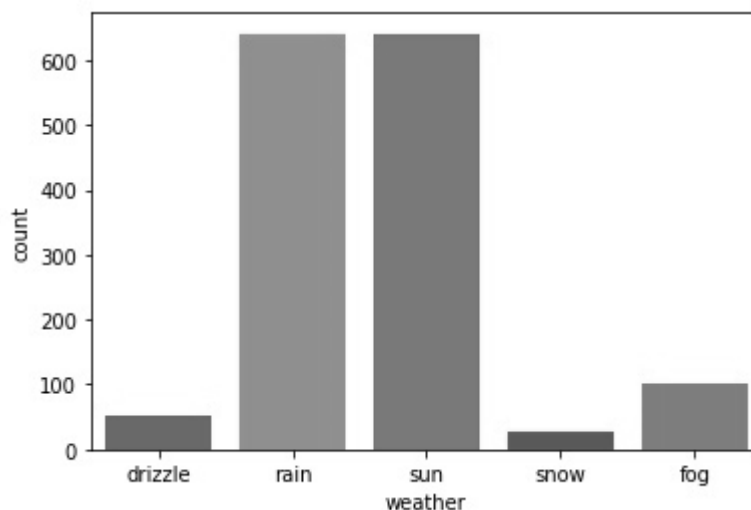


Рис. 3. График количества данных погоды [Авторская разработка]

На рис. 3. Изображен график, который показывает количество данных по каждому из видов погоды. Из данных следует, что 43% занимает дождь, 43% солнце, 4% морось, 2% снег, 8% туман.

С помощью вывода в таблицу были описаны полученные данные

	precipitation	temp_max	temp_min	wind
count	1461.000000	1461.000000	1461.000000	1461.000000
mean	3.029432	16.439083	8.234771	3.241136
std	6.680194	7.349758	5.023004	1.437825
min	0.000000	-1.600000	-7.100000	0.400000
25%	0.000000	10.600000	4.400000	2.200000
50%	0.000000	15.600000	8.300000	3.000000
75%	2.800000	22.200000	12.200000	4.000000
max	55.900000	35.600000	18.300000	9.500000

Рис. 4. Описание данных [Авторская разработка]

В данном блоке программы происходит построение графиков. На оси абсцисс обозначены значения каждого из признаков, а на оси ординат количество значений. Рис. 5.

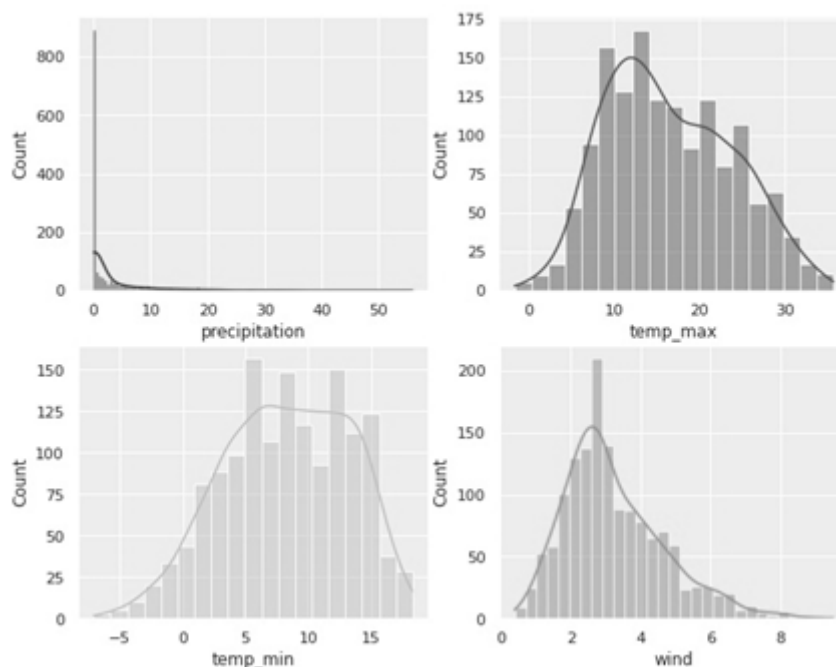


Рис. 5. График распределения данных [Авторская разработка]

Из приведённого выше рисунка ясно, что осадки и ветер имеют положительную асимметрию. Минимальная температура имеет отрицательную асимметрию. Это означает, что данные распределены неравномерно, что может отрицательно сказаться на обучении программы. С помощью графика скрипичной диаграммы можно понять асимметрию данных. (Рис. 6.)

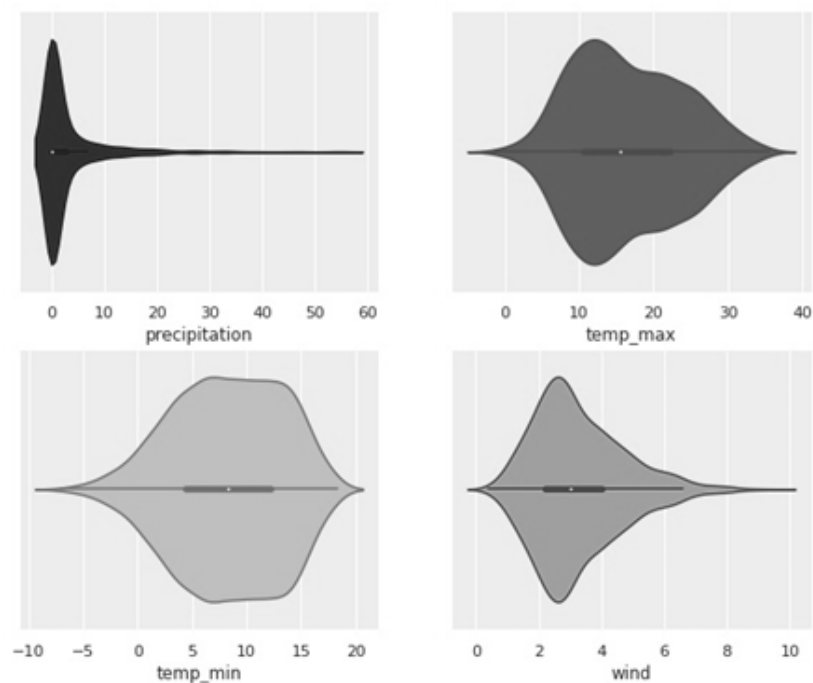


Рис. 6. Скрипичная диаграммы [Авторская разработка]

Далее был выведен график корреляции между признаками. Такой график необходим, что можно было наглядно увидеть зависимость между признаками.

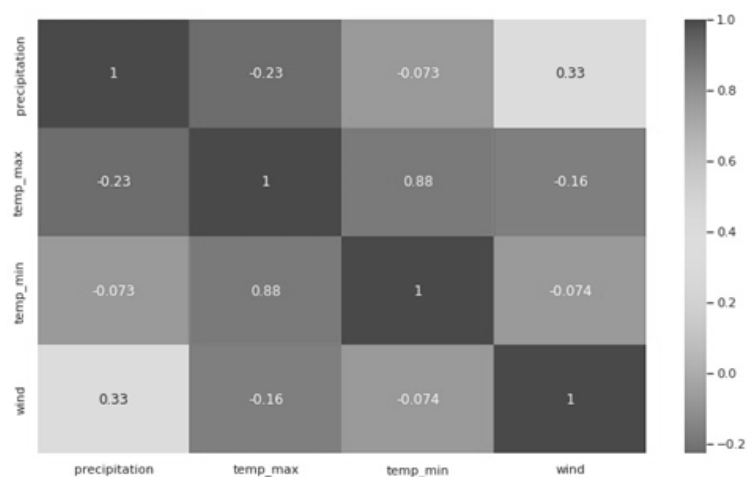


Рис. 7. График корреляции [Авторская разработка]



Следующим шагом была произведена обработка данных. Сначала данные были проверены на пропуски значений.

```
date          0
precipitation 0
temp_max      0
temp_min      0
wind          0
weather       0
dtype: int64
```

Рис. 8. Обработка данных [Авторская разработка]

Из рис. 8. Можно понять, что пропусков данных нет. Из базы данных был удалён столбец с датой, так как он не влияет на конечный результат.

Так как данные имеют выбросы (рис. 5.), то с помощью формулы, представленной ниже, данные будут отредактированы, то есть приведены к более равномерному виду.

```
Q1=df.quantile(0.25)
Q3=df.quantile(0.75)
IQR=Q3-Q1
df=df[~((df<(Q1-1.5*IQR))|(df>(Q3+1.5*IQR))).any(axis=1)]
```

Рис. 9. Формула, нормализующая данные [Авторская разработка]

Впоследствии была произведена обработка асимметрии данных (выпадение осадков, ветер) с помощью возведения под корень значений.

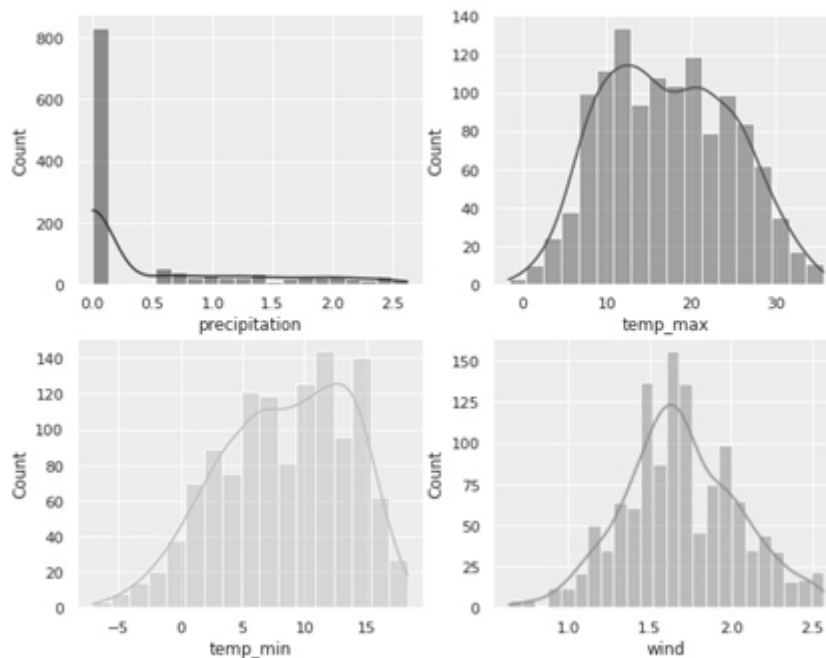


Рис. 10. Обновлённое распределение данных [Авторская разработка]

Следующим шагом было произведено масштабирование признака погоды с помощью метода кодирования меток. То есть значения погоды принимают числовой вид от 0 до 4.

Данные были разделены на тренировочные и тестовые.

Обучение модели было произведено с помощью алгоритма обучения extreme gradient boosting. После обучения модель показала 83% верных предсказаний на тестовых данных.

Таким образом, был получен полноценный ИИ, способный предсказывать погоду, исходя из начальных данных с достаточно большим процентом верных прогнозов. Уже сейчас программа может быть использована для решения различных задач.

**Библиографический список:**

1. Градиентный бустинг. [Электронный ресурс]. – Режим доступа – URL: <https://habr.com/ru/company/ods/blog/327250/> (Дата обращения 26.03.2022)
2. Kaggle. [Электронный ресурс]. – Режим доступа – URL: <https://www.kaggle.com> (25.03.2022)
3. Асимметрия и эксцесс эмпирического распределения. [Электронный ресурс]. – Режим доступа – URL: [http://www.mathprofi.ru/asimetriya\\_i\\_excess.html](http://www.mathprofi.ru/asimetriya_i_excess.html) (Дата обращения 27.03.2022)
4. Умная нормализация данных. [Электронный ресурс]. – Режим доступа – URL: <https://habr.com/ru/post/527334/> (Дата обращения 26.03.2022)
5. Способы обеспечения качества данных для машинного обучения. [Электронный ресурс]. – Режим доступа – URL: <https://habr.com/ru/post/588266/> (Дата обращения 26.03.2022)
6. Гид: алгоритмы машинного обучения и их типы. [Электронный ресурс]. – Режим доступа – URL: [https://www.sas.com/ru\\_ru/insights/articles/analytics/machine-learning-algorithms-guide.html](https://www.sas.com/ru_ru/insights/articles/analytics/machine-learning-algorithms-guide.html) (Дата обращения 25.03.2022)
7. Stepik. [Электронный ресурс]. – Режим доступа – URL: <https://welcome.stepik.org/ru> (Дата обращения 25.03.2022)
8. Введение в машинное обучение. [Электронный ресурс]. – Режим доступа – URL: <https://www.hse.ru/data/2017/05/14/1171296413/Григорий%20Сапунов%20—%20Введение%20в%20машинное%20обучение.pdf> (Дата обращения 23.03.2022)

9. 9 ключевых алгоритмов машинного обучения простым языком. [Электронный ресурс]. – Режим доступа – URL: <https://habr.com/ru/post/509472/> (Дата обращения 24.03.2022)
10. Алгоритмы машинного обучения [Электронный ресурс]. – Режим доступа – URL: <https://azure.microsoft.com/ru-ru/overview/machine-learning-algorithms/> (Дата обращения 24.03.2022)
11. Градиентный бустинг — просто о сложном [Электронный ресурс]. – Режим доступа – URL: <https://neurohive.io/ru/osnovy-data-science/gradientyj-busting/> (Дата обращения 25.03.2022)
12. Градиентный бустинг [Электронный ресурс]. – Режим доступа – URL: [https://ml-handbook.ru/chapters/grad\\_boost/intro](https://ml-handbook.ru/chapters/grad_boost/intro) (Дата обращения 25.03.2022)
13. Scikit-learn Machine Learning in Python бустинг [Электронный ресурс]. – Режим доступа – URL: <https://scikit-learn.org/stable/index.html> (Дата обращения 25.03.2022)
14. Методы статистической обработки результатов [Электронный ресурс]. – Режим доступа – URL: <https://u4isna5.ru/konspektlekcii/38-putvnauku/155-putvnauku212> (Дата обращения 24.03.2022)
15. Учебник по машинному обучению [Электронный ресурс]. – Режим доступа – URL: <https://ml-handbook.ru> (Дата обращения 24.03.2022)
16. Deepmind [Электронный ресурс]. – Режим доступа – URL: <https://www.deepmind.com> (Дата обращения 24.03.2022)
17. Астапов Р.Л., Мухамадеева Р.М. / АВТОМАТИЗАЦИЯ ПОДБОРА ПАРАМЕТРОВ МАШИННОГО ОБУЧЕНИЯ И ОБУЧЕНИЕ МОДЕЛИ МАШИННОГО ОБУЧЕНИЯ / Водяной О.А. // Актуальные научные исследования в современном мире. 2021. № 5-2 (73). С. 34-37. (Дата обращения 24.03.2022)

18. Галимов Р.Г. / ОСНОВЫ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ  
- ОБУЧЕНИЕ С УЧИТЕЛЕМ / Шелистов Д.А. // Аллея науки. 2017. Т.  
1. № 14. С. 810-817. (Дата обращения 23.03.2022)

*Оригинальность 80%*