

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РЕСПУБЛИКИ КАЗАХСТАН

Некоммерческое акционерное общество «Казахский национальный  
исследовательский технический университет имени К.И.Сатпаева»

Институт автоматизации и информационных технологий

Кафедра «Высшая математика и моделирование»

Сеитова Алия Айдаровна

Прогнозирование урожайности сельскохозяйственных культур с помощью  
машинного обучения

**ДИПЛОМНАЯ РАБОТА**

6B06103 - Математическое и компьютерное моделирование

Алматы 2023

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РЕСПУБЛИКИ КАЗАХСТАН

Некоммерческое акционерное общество «Казахский национальный исследовательский  
технический университет имени К.И.Сатпаева»

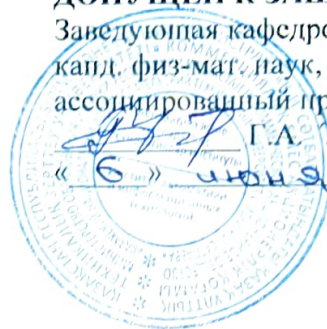
Институт автоматизации и информационных технологий

Кафедра «Высшая математика и моделирование»

**ДОПУЩЕН К ЗАЩИТЕ**

Заведующая кафедрой ВМиМ  
канд. физ-мат. наук,  
ассоциированный профессор

 Г.А. Тулшева  
« 5 » июня 2023 г.



**ДИПЛОМНАЯ РАБОТА**

На тему: «Прогнозирование урожайности сельскохозяйственных культур с помощью  
машинного обучения»

6B06103 – Математическое и компьютерное моделирование

Выполнил

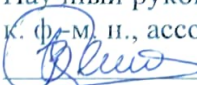
Сейтова.А.А

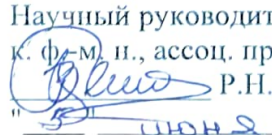
Рецензент  
PhD по направлению Математика  
ассоц. профессор АО КБТУ,  
Школа прикладной математики  
Д.О.Сарыбекова

  
« 5 » июня 2023 г.



Научный руководитель  
к. ф.-м. н., ассоц. профессор

 Р.Н.Зимин  
« 5 » июня 2023г.



Алматы 2023

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РЕСПУБЛИКИ КАЗАХСТАН

Некоммерческое акционерное общество «Казахский национальный  
исследовательский технический университет имени К.И.Сатпаева»

Институт автоматизации и информационных технологий

Кафедра «Высшей математики и моделирования»

**УТВЕРЖДАЮ**

Заведующий кафедрой ПИ  
канд. физ-мат. наук,  
ассоц. профессор

  
Гулешова.Г.А.  
« 8 » июня 2023 г.



**ЗАДАНИЕ**

**на выполнение дипломной работы**

Обучающемуся: Сеитова Алия Айдаровна

Тема: Прогнозирование урожайности сельскохозяйственных культур с  
помощью машинного обучения

Утверждена приказом проректора по академической работе: №\_408-П/Ө

от «23» ноября 2022 г.

Срок сдачи законченной работы

«8» июня 2023 г.

Исходные данные к дипломной работе: документация к «AGROTOOL v.3.5»,  
Python, Scikit-learn.

Краткое содержание дипломной работы:

*А) Материалы и методы;*

*Б) Результаты;*

*В) Выводы по результатам;*






*Г) Заключение.*

Рекомендуемая основная литература: техническая литература по «AGROTOOL  
v.3.5», прогнозированию урожайности, методам машинного обучения для  
прогнозирования, методу регрессии LASSO из 5 наименований.

**ГРАФИК**  
подготовки дипломной работы

Наименование разделов, перечень разрабатываемых вопросов	Сроки представления научному руководителю и консультантам	Примечание
Теория по использованным материалам и методам	27.01.23	Выполнено
Проведение эксперимента и получение результата	28.02.23	Выполнено
Анализ результата и определение корреляции переменных	21.03.23	Выполнено
Выводы, полученные из результатов и заключение.	18.05.23	Выполнено


**Подписи**  
консультантов и нормоконтролера на законченную дипломный проект  
с указанием относящихся к ним разделов проекта

Наименования разделов	Консультанты, И.О.Ф. (уч. степень, звание)	Дата подписания	Подпись
Теория по использованным материалам и методам	к. ф.-м. н., ассоц. профессор Зимин Р. Н.	28.02.23	
Проведение эксперимента и получение результата	к. ф.-м. н., ассоц. профессор Зимин Р. Н.	30.02.23	
Анализ результата и определение корреляции переменных	к. ф.-м. н., ассоц. профессор Зимин Р. Н.	25.03.23	
Выводы, полученные из результатов и заключение.	к. ф.-м. н., ассоц. профессор Зимин Р. Н.	25.05.23	
Нормоконтролер	к. ф.-м. н., ассоц. профессор Шатманов Ж. Ж.	5.05.23	

Научный руководитель

Задание принял к исполнению обучающийся

Дата

 Р. Н. Зимин

 А. А. Сеитова

« 10 » января 2023

**Университеттің жүйе администраторы мен Академиялық мәселелер департаменті  
директорының ұқсастық есебіне талдау хаттамасы**

Жүйе администраторы мен Академиялық мәселелер департаментінің директоры көрсетілген еңбекке қатысты дайындалған Плагиаттың алдын алу және анықтау жүйесінің толық ұқсастық есебімен танысқанын мәлімдейді:

**Автор: Сентова Алия Айдаровна**

**Тақырыбы: Прогнозирование урожайности сельскохозяйственных культур с помощью машинного обучения.doc**

**Жетекшісі: Решат Зимин**

**1-ұқсастық коэффициенті (30): 3.1**

**2-ұқсастық коэффициенті (5): 0.5**

**Дәйексөз (35): 1.3**

**Әріптерді ауыстыру: 3**

**Аралықтар: 0**

**Шағын кеңістіктер: 30**

**Ақ белгілер: 0**

**Ұқсастық есебін талдай отырып, Жүйе администраторы мен Академиялық мәселелер департаментінің директоры келесі шешімдерді мәлімдейді :**

Ғылыми еңбекте табылған ұқсастықтар плагиат болып есептелмейді. Осыған байланысты жұмыс өз бетінше жазылған болып санала отырып, қорғауға жіберіледі.

Осы жұмыстағы ұқсастықтар плагиат болып есептелмейді, бірақ олардың шамадан тыс көптігі еңбектің құндылығына және автордың ғылыми жұмысты өзі жазғанына қатысты күмән тудырады. Осыған байланысты ұқсастықтарды шектеу мақсатында жұмыс қайта өңдеуге жіберілсін.

Еңбекте анықталған ұқсастықтар жосықсыз және плагиаттың белгілері болып саналады немесе мәтіндері қасақана бұрмаланып плагиат белгілері жасырылған. Осыған байланысты жұмыс қорғауға жіберілмейді.

**Негіздеме:**

Күні 2.06.2023

Кафедра меңгерушісі



## Протокол

### о проверке на наличие неавторизованных заимствований (плагиата)

**Автор:** Сеитова Алия Айдаровна

**Соавтор (если имеется):**

**Тип работы:** Дипломный проект

**Название работы:** Прогнозирование урожайности сельскохозяйственных культур с помощью машинного обучения.doc

**Научный руководитель:** Решат Зимин

**Коэффициент Подобия 1:** 3.1

**Коэффициент Подобия 2:** 0.5

**Микропробелы:** 30

**Знаки из других алфавитов:** 3

**Интервалы:** 0

**Белые Знаки:** 0

**После проверки Отчета Подобия было сделано следующее заключение:**

- Заимствования, выявленные в работе, является законным и не является плагиатом. Уровень подобия не превышает допустимого предела. Таким образом работа независима и принимается.
- Заимствование не является плагиатом, но превышено пороговое значение уровня подобия. Таким образом работа возвращается на доработку.
- Выявлены заимствования и плагиат или преднамеренные текстовые искажения (манипуляции), как предполагаемые попытки укрытия плагиата, которые делают работу противоречащей требованиям приложения 5 приказа 595 МОН РК, закону об авторских и смежных правах РК, а также кодексу этики и процедурам. Таким образом работа не принимается.
- Обоснование:

Дата 2.06.2023

  
проверяющий эксперт

## Протокол

### о проверке на наличие неавторизованных заимствований (плагиата)

**Автор:** Сеитова Алия Айдаровна

**Соавтор (если имеется):**

**Тип работы:** Дипломный проект

**Название работы:** Прогнозирование урожайности сельскохозяйственных культур с помощью машинного обучения.doc

**Научный руководитель:** Решат Зимин

**Коэффициент Подобия 1:** 3.1

**Коэффициент Подобия 2:** 0.5

**Микропробелы:** 30

**Знаки из других алфавитов:** 3

**Интервалы:** 0

**Белые Знаки:** 0

**После проверки Отчета Подобия было сделано следующее заключение:**

Заимствования, выявленные в работе, является законным и не является плагиатом. Уровень подобия не превышает допустимого предела. Таким образом работа независима и принимается.

Заимствование не является плагиатом, но превышено пороговое значение уровня подобия. Таким образом работа возвращается на доработку.

Выявлены заимствования и плагиат или преднамеренные текстовые искажения (манипуляции), как предполагаемые попытки укрывтия плагиата, которые делают работу противоречащей требованиям приложения 5 приказа 595 МОН РК, закону об авторских и смежных правах РК, а также кодексу этики и процедурам. Таким образом работа не принимается.

Обоснование:

Дата 2.06.2023

Заведующий кафедрой



## РЕЦЕНЗИЯ

на дипломную работу  
Сеитовой Алии Айдаровны  
6B06103 – Математическое и компьютерное моделирование

На тему: «Прогнозирование урожайности сельскохозяйственных культур с помощью машинного обучения»

### ЗАМЕЧАНИЯ К РАБОТЕ

Перед дипломантом стояла задача прогнозирования урожайности ячменя с помощью методов машинного обучения. В качестве инструмента использовалась регрессии LASSO. Так же изучалось влияние изменения климата на производство сельскохозяйственной культуры.

Во введении определяется актуальность выбранной темы, цели и задачи эксперимента, объект и предмет исследования, методы сбора и анализа информации, обосновывается структура дипломной работы. Стоит уточнить, что для прогнозирования использовались синтетические данные полученные из программы AGROTOOL v. 3.5.

В первой главе данной дипломной работы рассматриваются исходные данные для эксперимента, область исследования, определения использованных инструментов и мотивация их выбора. Так в качестве тестирования была выбрана деревня Меньково (сельскохозяйственная деревня Ленинградской области). Данные по урожайности были собраны благодаря метеостанции Белогорка, которые были сгенерированы программным комплексом “AGROTOOL v.3.5.”.

Во второй главе рассматриваются результаты эксперимента, анализ и выводы. Выведены графики зависимости урожайности с погодными переменными и определены причины их корреляции. Метод регрессии LASSO показал свою применимость, о чем свидетельствуют ее значения коэффициента детерминации ( $R^2=0.987$ ), среднеквадратического отклонения ( $RMSE=0.113$ ) и средней абсолютной ошибки ( $MAE=0.081$ ). Это показывает применимость приведенного метода для прогнозирования урожайности. В работе так же рассматривалась корреляция между погодными условиями и урожайностью.

В заключении приведены общие выводы о проделанной работе.

В качестве замечания, не влияющего на общую оценку работы, следует отметить, что так как в качестве прогнозирования использовался лишь один метод машинного обучения, то сложно говорить о том, насколько он эффективен, так как не производится никакого сравнения с другими возможными методами. Тут следует говорить лишь о применимости данного метода в прогнозировании урожайности.

### Оценка работы

Учитывая актуальность дипломной работы, считаю, что дипломная работа заслуживает оценки «90%», а студент Сеитова А.А. при успешной защите – присвоения степени бакалавра по специальности 6B06103 – Математическое и компьютерное моделирование.

#### Рецензент

Ассоц. профессор АО КБТУ,  
Школа прикладной математики  
PhD по направлению Математика  
(доказательство наличия знаний)

\_\_\_\_\_ Л.О.Сарыбекова  
(подпись)

« 5 » \_\_\_\_\_ 2023 г.



## ОТЗЫВ

### НАУЧНОГО РУКОВОДИТЕЛЯ

На дипломную работу  
Сеитовой Алии Айдаровны  
6В06103 – Математическое и компьютерное моделирование

Тема: «Прогнозирование урожайности сельскохозяйственных культур с помощью машинного обучения»

Перед Сеитовой А.А. стояла следующая задача — прогнозирование урожайности ячменя с помощью методов машинного обучения.

Для построения прогноза использовались синтетические данные, полученные при помощи программы «AGROTOOL v. 3.5».

В качестве тестирования рассматриваемого метода прогноза была выбрана деревня Меньково. Данные по урожайности были сгенерированы программным комплексом «AGROTOOL v.3.5». Для построения прогноза урожайности использовался метод регрессии LASSO. Были поставлены численные эксперименты и проведен анализ полученных результатов, построены графики зависимости урожайности от погодных переменных и определены причины их корреляции. Метод регрессии LASSO показал вполне хорошую применимость к задаче прогнозирования урожайности.

Дипломная работа Сеитовой Алии Айдаровны состоит трех основных частей: теоретической части, основного результата и заключения.

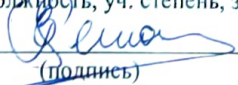
Считаю, что Сеитова А.А. справилась с поставленной ей задачей, ее дипломная работа соответствует всем выдвигаемым к таким работам требованиям. На основании вышесказанного считаю, что дипломная работа заслуживает оценки «90», а Сеитова А. А. — присуждения ей академической степени бакалавра по специальности 6В06103 — «Математическое и компьютерное моделирование».

**Научный руководитель**

Ассоц. профессор кафедры ВМиМ,

к.ф.-м.н.

(должность, уч. степень, звание)



Зимин Р.Н.

(подпись)

« 5 » июня 2023 г.

## АНДАТПА

Елдердегі ауыл шаруашылығы дақылдарының шығымдылығын болжау азық-түлікпен қамтамасыз етілмеу қаупін төмендетуі мүмкін.

Бұл қағазда Ленинград облысының Менково ауылында өсірілетін Белогорка метеостанциясынан алынған деректерді пайдалана отырып, арпаның өнімділігін болжау үшін машиналық оқыту әдісі қолданылады.

Жиналған деректер LASSO регрессиясын пайдалана отырып талданды.

Негізгі болжамдар ретінде ауа райының айнымалылары мен вегетативтік ағза массасы пайдаланылды. Үлгілер оқытылып, сынақтан өтті.

Эксперимент нәтижелері ауыл шаруашылығына байланысты шешімдер қабылдауда ассистент қызметін атқара алады.

## АННОТАЦИЯ

Прогнозирование урожайности в странах может привести к снижению риска отсутствия продовольственной безопасности.

В данной работе используется метод машинного обучения для прогнозирования урожайности ячменя, с использованием данных, полученных с метеостанции Белогорка, возделанном в селе Меньково Ленинградской области.

Собранные данные были проанализированы с помощью регрессии LASSO.

В качестве основных предикторов использовались погодные переменные и масса вегетативных органов. Модели прошли обучение и тестирование.

Результаты эксперимента могут служить помощником при принятии решений, связанных с сельским хозяйством.

## ABSTRACT

Forecasting crop yields in countries can reduce the risk of food insecurity.

This paper uses a machine learning method to predict barley yields, using data obtained from the Belogorka weather station cultivated in the village of Menkovo, Leningrad Region.

The collected data were analyzed using LASSO regression.

Weather variables and vegetative organ mass were used as the main predictors. The models have been trained and evaluated.

The results of the experiment can serve as an assistant in making decisions related to agriculture.

## СОДЕРЖАНИЕ

	Введение	7
1	Материалы и методы	9
1.1	Общие сведения	9
1.2	Полевой участок	9
1.3	Сельскохозяйственная культура возделования	10
1.4	Данные о почве	10
1.5	Данные о погоде	12
1.6	Метод исследования	14
1.7	Среда моделирования «AGROTOOL v.3.5»	15
1.8	Регрессия LASSO	18
1.9	Язык программирования Python	19
1.10	Оценка моделей	20
2	Результаты	21
2.1	Результаты прогноза	21
2.2	Важность переменных	26
2.3	Выводы результатов	29
2.4	Стадия выращивания сельскохозяйственных культур и климатические требования	30
2.5	Прогнозирование и производительность модели	30
	Заключение	31
	Список использованной литературы	32
	Приложение А. Техническое задание	34

## ВВЕДЕНИЕ

Сельское хозяйство является важной экономической частью большинства стран. Его зависимость от погодных условий делает результат непредсказуемым. И порой урожай не оправдывает ожиданий из-за обилия или недостатка тех или иных погодных явлений.

Изменчивость климата оказывает большое влияние на работу сельского хозяйства и может привести к таким серьезным проблемам, как голод или отсутствию продовольственной безопасности. Особенно в регионах, характеризующихся засухами или другими стихийными бедствиями, связанными с погодой. Климатические переменные, влияющие на урожайность, включают себя осадки, температуру воздуха, влажность, почвенные данные и солнечную радиацию [8]. Различные исследования показали, что климатические индексы как на глобальном, так и на региональном уровне влияют на урожайность и продовольственную безопасность [9]. Снижение урожайности может быть связано либо с высокой температурой, либо с обильными осадками.[10] Экстремальные температуры оказывают негативное влияние на урожай. Во всех сценариях изменение климата по-разному может повлиять на сельское хозяйство.

Климатические переменные могут быть похожими для конкретной местности, однако, потребности в погодных параметрах различаются от одной культуры к другой в зависимости от стадии их выращивания [8]. Это означает, что каждая культура имеет различный уровень устойчивости к атмосферным переменным.

Для решения проблем, связанных с изменением погодных условий для сельскохозяйственных культур, в различных исследованиях были предложены различные решения [11]. Исследование, проведенное для определения влияния экстремальных погодных условий на различные регионы Европы показало, что наиболее надежными погодными предикторами сельскохозяйственного производства являются количество осадков и температура воздуха [10]. Осадки и температура воздуха являются наиболее распространенными климатическими параметрами, используемыми во многих исследованиях. Однако, другие параметры, такие как солнечная радиация, влажность воздуха и почвы использовались для прогнозирования урожайности с применением моделей машинного обучения, одним из которых является регрессия LASSO [11].

В исследованиях, описанных в литературе, не приводилось соотношение вклада каждого параметра погоды в урожайность сельскохозяйственных культур. Кроме того, в этих исследованиях не указывалось, на какой стадии развития культуры более или менее необходим каждый погодный параметр. Знание вклада каждого климатического фактора и его порогового значения для хорошего производства имеет решающее значение для будущего

прогнозирования урожайности на основе мониторинга погоды с использованием моделей машинного обучения.

В этом исследовании был изучен метод регрессии LASSO для реализации в наших будущих исследованиях по прогнозированию урожайности. Данные по урожайности были собраны благодаря метеостанции Белогорка, которые были сгенерированы программным комплексом “AGROTOOL v.3.5.” [2,20].

В частности, основные этапы данной работы заключаются в следующем:

1 Определение корреляции между урожайностью и погодными параметрами (т. е. количеством осадков и температурой);

2 Определение важности каждого параметра погоды для производства сельскохозяйственных культур;

3 Определение эффективности регрессии LASSO для прогнозирования урожайности ячменя [2].

Остальная часть организована следующим образом. В разделе 1 представлено обсуждение материалов и методов исследования. Он включает теоретические данные об области и методе исследования, инструментах, использованных в работе и мотивацию их выбора. В разделе 2 представлены результаты с точки зрения производительности модели. Подробное обсуждение результатов и связь переменных между собой. Раздел 3 – заключение.

## **1. Материалы и методы**

### **1.1 Общие сведения**

Урожайность сельскохозяйственных культур – один из важнейших показателей эффективности аграрной системы местности и успешности их возделывания в конкретных условиях. Прогнозирование урожайности напрямую влияет на экономику страны и помогает заблаговременно планировать необходимые мероприятия для создания благоприятных условий выращивания и устранения возможных проблем для получения лучших результатов [17].

### **1.2 Полевой участок**

Место, выбранное для исследования – село Меньково. Меньково - населенный пункт в Ленинградской области Российской Федерации. Он расположен на северо-западе страны, на правом берегу реки Лужи. Географически Меньково находится, примерно, в 70 километрах к юго-западу от Санкт-Петербурга, а также вблизи границы с Эстонией.

Меньково находится в историческом регионе Ижора, который характеризуется уникальной природной средой. В окрестностях населенного пункта можно встретить лесные массивы, озера и реки, создающие живописный ландшафт. Здесь преобладает рельеф, характерный для Ленинградской области, с небольшими холмами и углублениями.

Меньково - это развивающийся район с разнообразной инфраструктурой. Здесь можно найти жилые зоны, коммерческие предприятия, магазины, школы, детские сады и другие объекты общественного назначения.

Ленинградская область является одним из крупнейших и наиболее развитых регионов Северо-Западного федерального округа Российской Федерации. Она граничит с Финляндией и Эстонией, а также с другими регионами страны, такими как Новгородская, Псковская и Вологодская области.

Природа Ленинградской области включает в себя озера, реки, леса и болота, а также природные заповедники и национальные парки, предлагающие множество возможностей для отдыха и экологического туризма.

В деревне находится опытная станция Агрофизического научно-исследовательского института (АФИ) Российской академии сельскохозяйственных наук. На 2014 год в деревне было учтено 242 домохозяйства.[18]

### **1.3 Сельскохозяйственная культура возделывания**

Ленинградская область имеет богатую структуру посевов по области. Кормовые культуры занимают наибольшую долю (71,0%) в структуре посевных площадей по всей Ленинградской области, яровой и озимый ячмень (10,0%), озимая и яровая пшеница (3,8%), овес (3,1%), картофель для промышленного выращивания (2,5%), яровая и озимая тритикале (1,3%) и овощи открытого грунта промышленного выращивания (1,3%).

Общие размеры всех посевных площадей Ленинградской области, по данным 1990 года, составляли 436,7 тыс. га.

По данным 2015 года регион занимает 57-ю строчку по размерам посевных площадей.

В качестве культуры исследования был выбран ячмень. Регион занимает 39-е место по сбору ячменя и 44-е место по размерам площадей, занятых под данной культурой [6].

#### **1.4 Данные о почве**

Почвенный покров Меньково и его окрестностей в Ленинградской области обычно относят к различным типам почв, в зависимости от конкретного района. На территории региона можно встретить следующие типы почв:

1 Подзолистые почвы: Широко распространены в лесистых районах Ленинградской области. Они образуются под влиянием хвойных и смешанных лесов. Подзолистые почвы имеют светлый верхний го-ризонт, богатый органическим веществом, и кислую или слабокис-лую реакцию.

2 Глеевые почвы: Характерны для равнинных и низменных районов, где преобладает речная депрессия или мелиоративные угодья. Глее-вые почвы образуются под влиянием водных процессов и имеют вы-сокую плодородность. Они хорошо удерживают влагу и обладают умеренной кислотностью.

3 Мезотрофные и дерново-подзолистые почвы: встречаются на рых-лых отложениях песчаников и глинистых песчаников.

4 Болотные почвы: Характерны для болотистых и заболоченных рай-онов Ленинградской области. Они образуются в условиях высокой влажности и слабого доступа кислорода. Болотные почвы обладают низкой плодородностью и имеют кислую реакцию.

Эти почвы имеют среднюю плодородность и хорошие водоудерживающие свойства. Они широко используются в сельском хозяйстве для земледелия и садоводства.

Важно отметить, что конкретные типы почв и их свойства могут варь-роваться в зависимости от местных условий, климата и геологических особенностей. Например, во влажных местностях и лиственных лесах образовались дерново-подзолистые почвы [4].

На Рисунке 1.3.1 и 1.3.2 ниже показаны административная граница села Меньково и его расстояние до метеостанции района – Белогорка соответственно. Расстояние между объектами составляет 15,7 километра [3].

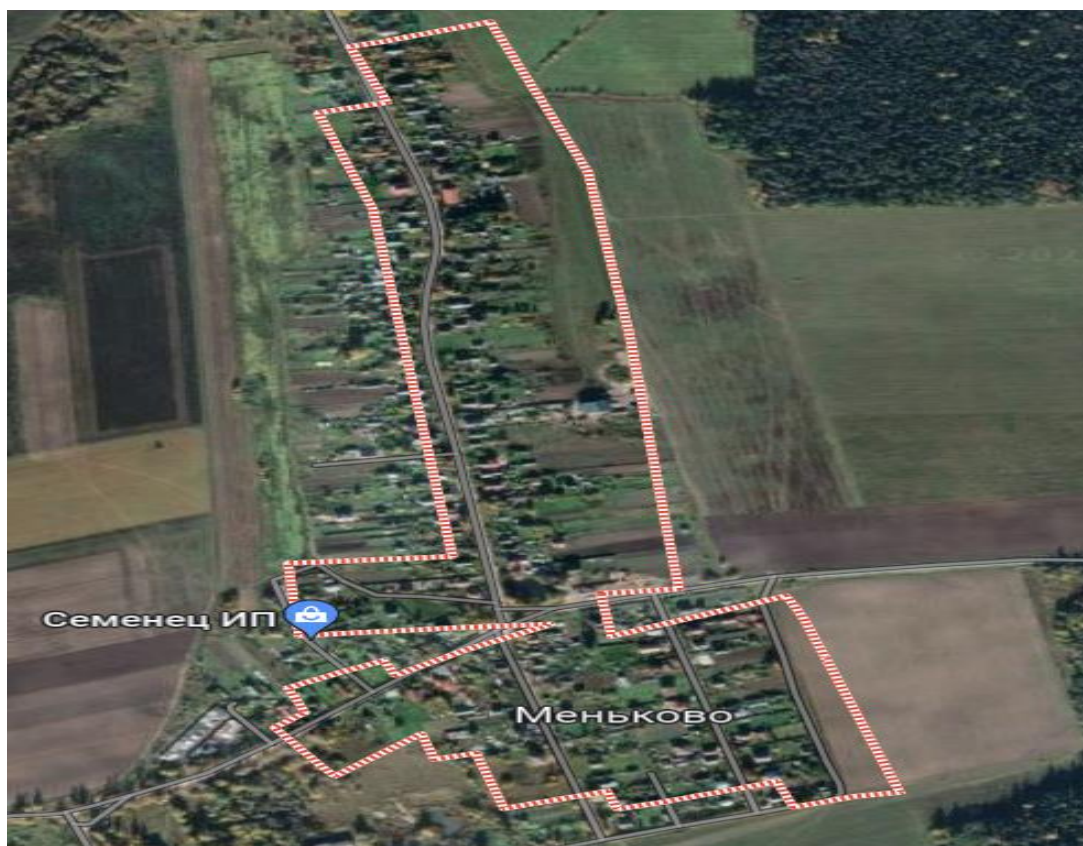


Рисунок 1.3.1 - Административная граница села Меньково.

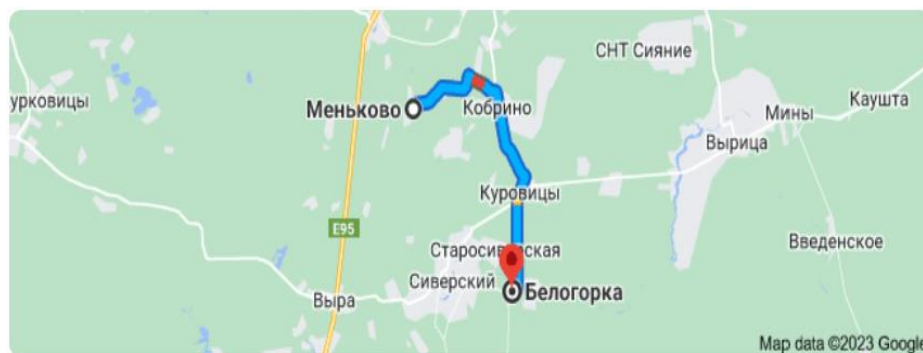


Рисунок 1.3.2 - Расстояние села Меньково от метеостанции Белогорка (15,7 км)



## 1.5 Данные о погоде

Основными погодными параметрами, влияющими на развитие сельскохозяйственных культур, являются осадки, температура и влажность воздуха, солнечная радиация скорость ветра.

В эксперименте по прогнозированию урожайности использовались погодные и вегетативные показатели за 1988 год, доступные в операционной базе данных программного комплекса “AGROTOOL v.3.5”. Данные были получены метеостанцией Белогорка, Ленинградской области.

Поскольку сельскохозяйственные культуры имеют различные потребности в воде в зависимости от фазы их развития, месячное общее количество осадков было рассчитано суммарно, основываясь на ежедневных наблюдениях.

Для прогнозирования урожайности ячменя использовались среднемесячные значения температуры, поскольку требования к температуре различаются также в зависимости от фазы развития урожая сельскохозяйственной культуры, аналогично осадкам.

Ниже описана температура, необходимая для получения оптимального урожая ячменя. Семена ячменя прорастают при температуре 1-2°C. Восход жизнеспособных ростков может проявиться при 4-5°C. Однако, оптимальной температурой для прорастания ячменя 20-23°C. Ячмень способен перенести заморозки до -8°C. [4]

На рисунке 1.5.1 показаны значения данных о погоде, а также их расшифровка на рисунке 1.5.2.

	Number	DATE	TMIN	TMAX	HUM	PREC	WIND	Kex	TSURF
0	0	1988-01-01	-15.5	-7.0	76.0	0.0	4.0	0.359	NaN
1	1	1988-01-02	-6.3	1.8	90.0	3.9	4.0	0.254	NaN
2	2	1988-01-03	1.0	4.9	86.0	5.5	4.0	0.254	NaN
3	3	1988-01-04	3.3	4.0	85.0	0.0	4.0	0.359	NaN
4	4	1988-01-05	0.2	3.9	88.0	1.4	4.0	0.254	NaN
...	...	...	...	...	...	...	...	...	...
726	726	1989-12-28	-3.1	-2.2	92.0	0.0	4.0	0.284	NaN
727	727	1989-12-29	-3.0	-0.8	93.0	0.0	4.0	0.284	NaN
728	728	1989-12-30	-2.3	-0.7	91.0	0.0	4.0	0.284	NaN
729	729	1989-12-31	-2.7	-1.4	79.0	0.2	4.0	0.252	NaN
730	12000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

731 rows × 9 columns

Рисунок 1.5.1 - Данные о погоде, использованные для прогноза.

Закладка «WEATHER» - Данные о погоде

№	Имя поля	Описание	Ед. изм.	Примечание
1	Number	Номер записи по порядку		
2	DATE	Дата		
3	TMIN	Минимальная температура воздуха за сутки	град. С	
4	TMAX	Максимальная температура воздуха за сутки	град. С	
5	HUM	Минимальная относительная влажность воздуха за сутки	%	Допустимо использовать среднюю влажность
6	PREC	Суточная сумма осадков	мм	
7	WIND	Среднесуточная скорость ветра	м/с	При отсутствии данных – допустимо задавать постоянное значение 4
8	Kex	Коэффициент ослабления приходящей коротковолновой радиации		Представляет собой отношение суточной суммы фактически поступившей к ЗЕМНОЙ ПОВЕРХНОСТИ общей (прямой и рассеянной) солнечной радиации к суточной сумме радиации, поступающей в данный день на данной широте на горизонтальную площадку НА ВЕРХНЕЙ ГРАНИЦЕ АТМОСФЕРЫ. Вычисляется явно (по актинометрическим измерениям) или оценивается по облачности или по показаниям гелиографа. Оценочное минимальное значение – 0.22 (абсолютно облачный день). Оценочное максимальное значение – 0.76 (абсолютно безоблачный день).
9	TSURF	Температура поверхности почвы	град. С	Допустимо не задавать. В этом случае будет рассчитываться в модели

Рисунок 1.5.2 - Расшифровка таблицы «Данные о погоде».

## 1.6 Метод исследования

Существует множество методов прогнозирования урожайности зерновых культур, но их практическое применение не всегда учитывает структуру посевных площадей и сортовую структуру, что могло бы обеспечить максимальный и точный эффект прогноза. Это объясняется недостатками существующей методологии краткосрочного и среднесрочного прогнозирования.

Одним из основных недостатков современной методологии прогнозирования является предположение об идеальном развитии зернового хозяйства и ежегодном приросте урожайности относительно базового уровня. Однако, фактическая динамика урожайности сельскохозяйственных культур в период с 1990 по 2015 годы свидетельствует о несоответствии этому предположению. Ежегодная объем посевных площадей уменьшается, соответственно, и количество получаемого урожая не имеет большого прироста.

Основная цель классификации методов прогнозирования урожайности заключается в выборе моделей, которые обеспечивают высокую точность прогнозов. Существует четыре основные группы методов, из которых выделяются: экстраполяция, моделирование, расчет урожайности на основе прироста от применения удобрений, а также экспертные оценки и аналогии [12,14].

В данной работе был использован метод прогнозирования урожайности – моделирование. В качестве среды моделирования был использован

программный комплекс “AGROTOOL v.3.5”, который имеет динамическую среду моделирования, которая позволяет получить данные, не описанные в операционной базе данных (ОБД). Она способна генерировать ежедневные показатели, используя на вход переменные, доступные в ОБД. То есть создать имитацию реального полевого опыта. Подробнее о данной системе динамического программирования будет описано ниже в разделе 1.7 [2].

Для получения максимальной эффективности прогноза урожайности ячменя среда динамического программирования “AGROTOOL v.3.5” была использована в сочетании с машинным обучением.

Различные исследования показали, что машинное обучение является важным инструментом поддержки принятия решений для прогнозирования урожайности [12,14,]. Машинное обучение — это технология, которая может помочь фермерам сократить свои сельскохозяйственные потери, предлагая подробные рекомендации и информацию по урожаю.

В данном эксперименте для прогнозирования урожайности ячменя используется метод регрессии LASSO. Эти подходы были выбраны на основе числового характера прогноза и размера набора данных. Подробнее о регрессии LASSO и мотивации выбора именно данной модели, в качестве основной для проведения эксперимента будет описано в разделе 1.8.

### **1.7 Среда моделирования «AGROTOOL v. 3.5»**

Для получения необходимых данных с метеостанции Белогорка и различных почвенных и технических показателей используется среда динамического моделирования «AGROTOOL v. 3.5». В данном программном комплексе существуют стационарная и оперативная базы данных.

Программный комплекс «AGROTOOL v. 3.5» предназначен для проведения модельных динамических расчетов продукционного процесса сельскохозяйственных культур. Структура данного комплекса включает в себя два независимых исполняемых программных компонента и реляционную базу данных. На рисунке 1.7.1 представлена схема упрощенной концептуальной модели базы данных.

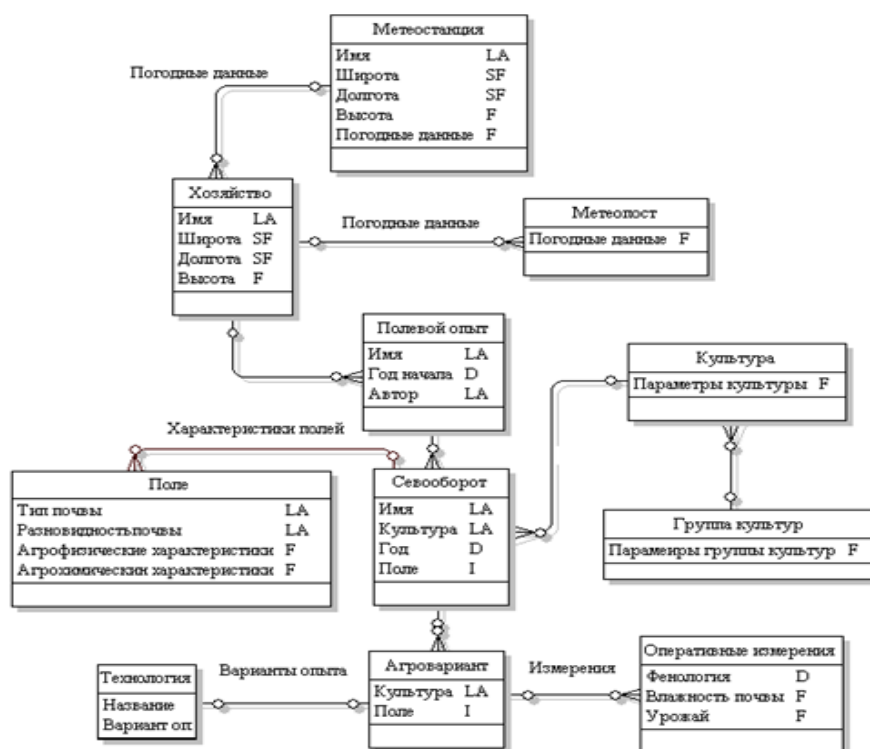


Рисунок 1.7.1 - Упрощенная концептуальная модель базы данных.

База данных полевого опыта и параметры модели состоят из набора нормализованных таблиц данных, которые содержат информацию двух типов. К первому типу относятся фактические сведения о характеристиках условий выращивания растений в наблюдаемых производственных посевах – погодные данные, параметры почв, структура севооборота, описание используемых технологий для посева и сбора, а также результаты обследований посевов и данные об урожайности в конкретные годы в конкретной местности. Эта информационная составляющая. Ко второму типу относятся сведения, имеющие специализированную направленность и относящиеся непосредственно к информационному обеспечению задач моделирования, например, специфические для модели параметры сельскохозяйственных культур и сортов.

Основные функциональные возможности, предоставляемые моделью:

- 1 Пошаговый расчет динамики агроэкосистемы с возможностью контроля ее состояния на каждом временном шаге.
- 2 Полный расчет динамики продукционного процесса на всем протяжении вегетационного периода (от посева до уборки).
- 3 Учет как predetermined планов технологических воздействий (жестко назначенные даты и нормы), так и исследование гибких условных алгоритмов агротехнических мероприятий, основанных на триггерных схемах их применения (запуск по достижению тех или иных условий, определяемых состоянием агроэкосистемы).

- 4 Использование единой алгоритмической модели продукционного процесса третьего уровня продуктивности для описания различных типов сельскохозяйственных культур. Переключение от одного типа культур к другому осуществляется исключительно выбором набора параметров в оперативной базе данных (ОБД). Моделирование в единой среде яровых и озимых злаковых культур, корнеплодов и многолетних трав.
- 5 Табличная и графическая визуализация результатов расчета в полностью настраиваемой и управляемой в RunTime режиме графической среде.
- 6 Оформление внутренней бизнес логики модели в виде набора функций, обеспечивающих ее базовый функционал и зарегистрированных в рамках выделенной динамически подключаемой библиотеки. Последнее дает возможность использовать модель из произвольных внешних оболочек и средств моделирования, предоставляемых сторонними разработчиками.

Операционная база данных (ОБД) представляет собой файл с расширением xls, содержащий стандартный набор именованных диапазонов. Структура данных этого файла соответствует структуре DataSet`а, принимаемого программой, и представляет собой необходимый комплект данных, подаваемых на вход алгоритму модели. ОБД формируется в результате выполнения специализированного интерфейса [2].

ОБД состоит из следующих основных разделов: Place, Weather, Air, Culture, Soil, Technology, InitialState, CommonDate

На рисунке 1.7.2 представлены таблицы данных разделов ОБД.

**4. Yield\_Growth\_Functions**

A1 e	A2 e	Type e	X0 e	Xmax e	Ymax e
------	------	--------	------	--------	--------

**5. Phenology\_Parameter**

Number	PhaseName	Tmin	Bmax	Pmin	Pstr1	Pstr2	Biopor
--------	-----------	------	------	------	-------	-------	--------

**6. Biom Regression\_Dependence**

Y0	Y5	Y10
----	----	-----

**7. LAI Regression\_Dependence**

Kl y 0	Kl y 1	Kl y 2
--------	--------	--------

**8. Photosynthesis and Productivity**

Ph_max	alfa	Cexpn	ResMes	Rx	Crst1	T0
--------	------	-------	--------	----	-------	----

Параметры из группы культур

**1. Plant\_Architect**

Bm_wet	Bmdry	Rcond	Clst	Hcmax	HmaxR	HminR	Cleaf	Rf	QU1	Vmax	KNDef	Nfl	Nyield
--------	-------	-------	------	-------	-------	-------	-------	----	-----	------	-------	-----	--------

**2. ROOT\_Growth\_Parameters**

RRM	RCVF	Nroot	CR	GR	RBini
-----	------	-------	----	----	-------

**3. LEAF\_Growth\_Parameters**

LRM	LCVF	Nleaf 1	Nleaf 2	CL	LBini	X L	Y L
-----	------	---------	---------	----	-------	-----	-----

**4. Yield\_Growth\_Parameters**

ERM	ECVF	Near	Lreut Coef	Sreut Coef
-----	------	------	------------	------------

**5. STEAM\_Growth\_Parameters**

SRM	SCVF	Nstem	CS	GS	X S	Y S
-----	------	-------	----	----	-----	-----

**Place (данные о местности)****1. Метеостанция**

NAME	LATITUDE	LONGITUDE	ALTITUDE	H FLUG	H METEO	MAX RSHPLANT
------	----------	-----------	----------	--------	---------	--------------

**2. Место**

NAME	LATITUDE	LONGITUDE	ALTITUDE	CO2Concentration
------	----------	-----------	----------	------------------

**3. Участок**

REL X	REL Y	REL Z	SLOPE AZIMUTH	SLOPE STEEP
-------	-------	-------	---------------	-------------

**Weather (данные о погоде)**

Number	DATE	TMIN	TMAX	HUM	PREC	WIND	Kex
--------	------	------	------	-----	------	------	-----

**Culture (данные о культуре)**

Параметры конкретной культуры

**1. Common Parameters**

Name
------

**2. ROOT\_Growth\_Functions**

A1 r	A2 r	Type r	X0 r	Xmax r	Ymax r
------	------	--------	------	--------	--------

**3. LEAF\_Growth\_Functions**

A1 l	A2 l	Type l	X0 l	Xmax l	Ymax l
------	------	--------	------	--------	--------

**InitialState (начальные условия)****1. Initial State**

WaterStorage
--------------

**CommonData (Описание dataset`а)****1. CommonData**

DataSet Type	DataSet Name
--------------	--------------

<i>Soil</i> (данные о почве)									
<b>1. Soil Surface</b>									
alb min	alb max	grayness							
<b>2. Physical Properties</b>									
Number	LowBound	RoSoil	RoSSP						
<b>3. Hydro Properties</b>									
Number	LowBound	TetMin	Wz	Fc	TetMax	al1	bl1	Kf	bl2
 <i>Technology</i> (данные о технологии)									
<b>1. Planting</b>									
Date	Doze	Descr	Depth	Fertility					
<b>2. Irrigation Regime</b>									
isTriggered	Kind	MinLimit	MaxLimit						
<b>3. Irrigation Actions</b>									
Number	Date	Doze	Descr	Kind					
<b>4. Fertilization Regime</b>									
isTriggered	Kind	MinLimit	MaxLimit						
<b>5. Fertilization Actions</b>									
Number	Date	Doze	Descr	Kind					
<b>6. Harvesting Regime</b>									
isTriggered	Kind	MaxVegLength	CtPhNum	YieldType					
<b>7. Harvesting Actions</b>									
Number	Date	Doze	Descr	Kind					

Рисунок 1.7.2 - Таблицы данных разделов ОБД.

## 1.8 Регрессия LASSO

Лассо (Least Absolute Shrinkage and Selection Operator) — это метод регрессии, который используется для выбора и оценки значимых переменных в модели регрессии. Он основан на добавлении штрафа L1-нормы (абсолютная сумма коэффициентов) к функции потерь, что позволяет выполнить сжатие коэффициентов и автоматически отобрать наиболее важные признаки. Именно поэтому для проведения эксперимента был выбран данный метод. [19]

Процесс регрессии Lasso состоит в минимизации следующей функции потерь:

$$\min \|Y - X\beta\|^2 + \alpha * \|\beta\|_1 \quad (1)$$

Где:

Y представляет собой вектор зависимой переменной,

X - матрица признаков,

$\beta$  - вектор коэффициентов регрессии,

$\|\dots\|^2$  обозначает квадрат нормы,

$\|\beta\|_1$  обозначает L1-норму коэффициентов,

$\alpha$  - параметр регуляризации, который контролирует силу сжатия коэффициентов.

Главной особенностью Lasso является то, что он способен устанавливать некоторые коэффициенты регрессии точно равными нулю. Это делает его

полезным инструментом для отбора признаков и создания более интерпретируемых моделей. Однако следует отметить, что выбор параметра регуляризации  $\alpha$  является важным шагом при использовании регрессии Lasso. Значение  $\alpha$  определяет компромисс между точностью модели и степенью сжатия коэффициентов. Если  $\alpha$  слишком большое, модель может быть слишком простой и недостаточно гибкой, а если  $\alpha$  слишком маленькое, модель может переобучиться и стать чувствительной к шуму в данных.

Выбор оптимального значения  $\alpha$  может осуществляться с помощью инструмента для автоматического подбора параметров для моделей машинного обучения - GridSearchCV.

GridSearchCV позволяет находить наилучшие параметры путем обычного перебора. Он принимает на вход саму модель и разные значения гиперпараметров. Далее рассчитывает ошибку для каждого из них и выбирает сочетание, при котором ошибка минимальна.

Регрессия Lasso является мощным инструментом для регуляризации моделей регрессии и отбора признаков. Она может быть применена в различных областях, включая машинное обучение, статистику и эконометрику, и позволяет создавать более простые и интерпретируемые модели без потери предсказательной способности.

## 1.9 Язык программирования Python

Для проведения эксперимента в качестве языка программирования был выбран язык Python. Он является одним из наиболее популярных языков программирования в области машинного обучения и анализа данных. Он просто в понимании, использовании и имеет широкий спектр инструментов и библиотек для решения задач машинного обучения. [20]

Для прогнозирования урожайности ячменя были использованы следующие инструменты:

- 1 Pandas — это мощная библиотека для анализа данных. Она позволяет
- 2 загружать, обрабатывать и анализировать структурированные данные, такие как исторические данные по урожайности и погоде
- 3 NumPy — это основная библиотека для работы с многомерными массивами данных в Python. Она предоставляет мощные функции и операции для эффективной обработки и вычислений с массивами.
- 4 Scikit-learn предоставляет широкий набор алгоритмов машинного обучения, включая регрессию LASSO и многое другое. Используется для построения модели прогнозирования урожайности.
- 5 Matplotlib — это библиотека визуализации данных. Она используется для отображения, чтобы визуально представить результаты прогнозирования урожайности.



## 1.10 Оценка моделей

Оценка моделей является важным шагом в машинном обучении и позволяет определить, насколько хорошо модель работает на тестовых данных или в реальных условиях. Существует множество показателей качества, которые можно использовать для оценки модели в зависимости от типа задачи. [20,14]

В задачах регрессии распространенными метриками для оценки качества являются MAE, RMSE и R-квадрат ( $R^2$ ).

Разница исходного и прогнозируемого значений представлена MAE, которая усредняет абсолютную разницу по всему набору данных.

$$MAE = \frac{1}{H} \sum_{i=1}^H |y_i - \hat{y}| \quad (2)$$

RMSE означает остаточную среднеквадратичную ошибку. Отделение точки данных от линии регрессии измеряется остатками, а дисперсия этих остатков измеряется RMSE. Другими словами, он показывает, насколько близко данные сосредоточены на линии наилучшего соответствия.

$$RMSE = \sqrt{\frac{1}{H} \sum_{i=1}^H |y_i - \hat{y}|^2} \quad (3)$$

R-квадрат (Коэффициент детерминации) является мерой того, насколько хорошо значения соответствуют друг другу по отношению к начальным значениям. Проценты представлены значениями от 0 до 1. Чем лучше модель, тем выше значение.

$$R^2 = 1 - \frac{\sum(y_i - \hat{y})^2}{\sum(y_i - \bar{y})^2} \quad (4)$$

где в уравнениях (2) – (4)  $y$  — прогнозируемое значение, а  $\hat{y}$  — это реальное значение.

## 2. Результаты

### 2.1 Результаты прогноза

Для первого численного эксперимента была использована таблица, сгенерированная программой “AGROTOOL v.3.5.”, с описанием суммарных осадков, температуры, вегетативными показателями, суммарной радиацией и параметров почвы. Также, в дополнение к основным показателям, были использованы параметры, значения которых не было указано в таблице. Это такие параметры, как удельная влажность воздуха и насыщения листа, суммарная транспирация, длина солнечного дня в часах. Для их вычисления использовалась программа “AGROTOOL v.3.5.”, которая принимала суммарные погодные, почвенные и вегетативные параметры в качестве входных для вычисления пошагового расчета удельной влажности воздуха и насыщения листа, суммарной транспирации и длины солнечного дня.

На рисунках 2.1.1 и 2.1.2 можно увидеть использованную таблицу для численного эксперимента и её расшифровку её параметров соответственно.

1 final_data																
NUMBER	DATE	Ebiom#1	Lbiom#1	Sbiom#1	SumBiom#1	Rbiom#1	PhTime#1	IPhase#4	LAI#1	...	SumNLow#1	SumNPlant#1	ExtNPool#1	WP		
0	1	2010-05-06	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	посев	0.000000	...	4.523981e-08	0.000000	0	173	
1	2	2010-05-07	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	посев	0.000000	...	9.081686e-08	0.000000	0	173	
2	3	2010-05-08	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	посев	0.000000	...	1.380244e-07	0.000000	0	173	
3	4	2010-05-09	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	посев	0.000000	...	1.806789e-07	0.000000	0	173	
4	5	2010-05-10	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	посев	0.000000	...	2.378862e-02	0.000000	0	136	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
91	92	2010-08-05	18.115277	9.421787	33.555870	61.092934	-33.167597	1.892602	молочная спелость	0.224863	...	1.260450e-01	89.817664	0	423	
92	93	2010-08-06	18.597556	9.258877	33.373349	61.229782	-32.880851	1.928089	молочная спелость	0.152114	...	1.260467e-01	89.817664	0	423	
93	94	2010-08-07	19.169582	9.066504	33.156876	61.392962	-32.539265	1.961338	воковая спелость	0.077895	...	1.260482e-01	89.817664	0	423	
94	95	2010-08-08	19.698665	8.889520	32.956673	61.544857	-32.221672	1.995692	воковая спелость	0.007917	...	1.260494e-01	89.817664	0	423	
95	96	2010-08-09	20.200531	8.722482	32.766790	61.689803	-31.918942	2.030316	полная спелость	-0.058763	...	1.260504e-01	89.817664	0	423	

96 rows x 38 columns

Рисунок 2.1.1 - Таблица «Результаты», использованная для первого численного эксперимента.

**Закладка «RESULTS» - Результаты**

Формируется моделью по результатам расчета				
№	Имя поля	Описание	Ед. изм.	Примечание
1	Number	Порядковый номер измерения		
2	DATE	Дата		
3	Lbiom#1	Биомасса листьев	ц/га	Символы «#<число>» в имени поля определяют тип графического представления соответствующей серии данных (1-линия, 2-столбчатая диаграмма и т.д.)
4	Sbiom#1	Биомасса стеблей	ц/га	
5	Ebiom#1	Биомасса генеративного органа	ц/га	
6	SumBiom#1	Суммарная надземная биомасса	ц/га	
7	Rbiom#1	Биомасса корней	ц/га	
8	PhTime#1	Физиологическое время	б/р	
9	IPhase#4	Наименование фазы		
10	LAI#1	Листовой индекс	м <sup>2</sup> /м <sup>2</sup>	
11	PrimAss#1	Накопленные ассимиляты	г/см <sup>2</sup>	
12	Tave#1	Среднесуточная температура воздуха град. С	Град. С	
13	SumRad#1	Суммарная радиация	Дж/(м <sup>2</sup> *сут)	
14	Pres#2	Осадки	мм	
15	Hroot#1	Глубина проникновения корней	см	
16	WCsoil#1	Влагозапас в метровом слое	см	
17	Eplant#1	Транспирация	мм/сут	
18	Esoil#1	Испарение с поверхности почвы	мм/сут	

19	SumPr/SumE#1	Отношение суммарных осадков к суммарной транспирации	б/р	
20	Sum_Snow#1	Высота снежного покрова (для озимых культур)	см	
21	Watering#1	Полив	см	
22	BioTime#1	Сумма активных температур	град. С	
23	SSPool#1	Пул запасного углерода (крахмал)	г/см <sup>2</sup>	
24	SNPool#1	Пул запасного азота (нитраты)	г/см <sup>2</sup>	
25	CRS#1	Доля первичных ассимилятов, идущих в корень		
26	Yield#1	Хозяйственный урожай	ц/га	
27	SumWat#1	Суммарный полив ??? убрали ???	см	
28	Nwat#1	Число поливов ??? убрали ???		
29	TotalNmin	Общее содержание азота в минеральных соединениях в почве	кг/га	
30	SumNUpt	Суммарный поглощенный азот	кг/га	
31	SumNDen	Суммарная денитрификация	кг/га	
32	SumNLow	Суммарное количество азота, выносимого за пределы метрового слоя почвы	кг/га	
33	SumNPlant	Суммарное количество азота в растении	кг/га	
34	ExtNPool	Азот на поверхности листьев после азотной подкормки	кг/га	
35	WPotSeed	Водный потенциал почвы на глубине заделки семян	см. вод. ст.	
36	TPotSeed	Температура почвы на глубине заделки семян	град. С	

Рисунок 2.1.2 - Расшифровка таблицы «Результаты».

Как обсуждалось выше, в разделе 1.10, производительность моделей оценивалась с использованием трех показателей: MAE (средняя абсолютная ошибка), RMSE (среднеквадратическое отклонение), R-квадрат (коэффициент детерминации).

Для определения наилучшего параметра регуляризации была проведена перекрестная проверка с использованием инструмента для автоматического подбора GridSearchCV проводилась среди 5 разных значений: [0.001, 0.01, 0.1, 0.5, 1]. Инструмент выбрал наиболее подходящим параметром регуляризации  $\alpha=0,001$ .

Выбор именно данных значений для перекрестной проверки был обусловлен изучением статей на тему параметра регуляризации [20]. Согласно наблюдениям, данный параметр чаще всего встречается в значениях меньше единицы. Так как слишком большое значение параметра регуляризации может привести к недообучению, занулив большинство важных параметров, а слишком малое приведет к переобучению, как обсуждалось ранее в разделе 1.8.

Результаты модели с параметром  $\alpha=0,001$ , выбранной после перекрестной проверки, показала следующие результаты (Таблицы 1):

Таблица 1 - Результат модели с параметром регуляризации, выбранным в результате перекрестной проверки.

Оценщик моделей	Результаты	
R <sup>2</sup> (коэффициент детерминации)	0.999 ( <i>Train</i> )	0.987 ( <i>Test</i> )
RMSE (среднеквадратическое отклонение)	0.11338038219738471	
MAE (средняя абсолютная ошибка)	0.07974589008840262	

Результаты, представленные в приведенной выше таблице, показывают, что результат прогноза модели регрессии LASSO соответствует данным об урожайности ячменя, на что указывает соответствующее значение R<sup>2</sup> (коэффициент детерминации), которое на тренировочных и тестовых данных показала результат 0.999 и 0.987 соответственно.

MAE (средняя абсолютная ошибка) между прогнозируемыми значениями и фактическими значениями для ячменя составляет 0.0797. А RMSE (среднеквадратическое отклонение) между прогнозируемой и фактической урожайностью составила 0.1134. Основываясь на полученные результаты, можно сделать вывод, что регрессия LASSO показывает прогноз урожайности ячменя.

Для визуализации результата был создан график с помощью библиотеки “matplotlib”, описанной в разделе 1.9 который позволяет увидеть отображение результатов в графической форме (Рисунок 2.1.3).

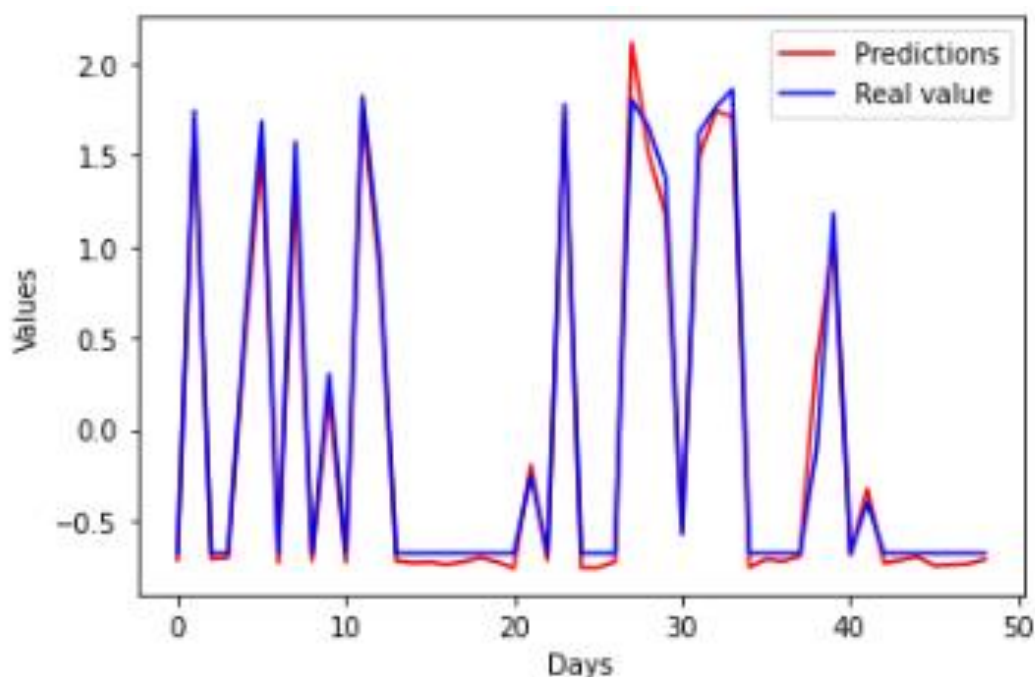


Рисунок 2.1.3 - Визуализация спрогнозированных и реальных значений урожайности ячменя (Красная линия - спрогнозированные значения, синяя линия – реальные значения)

Судя по данным из первого численного эксперимента, очевидно, что наша модель прогнозирует почти с идеальным результатом. Учитывая, что модель принимала на вход тренировочные и тестовые данные в соотношении 50 на 50, то есть, половину данных определяла для тестовых значений, а остальную для тренировочных, было принято решение протестировать её на новых данных, но, при этом, уменьшив количество тренировочных данных до 15 процентов, а для тестовых выделить оставшиеся 85 процентов.

В качестве предикторов на вход подаются новые данные об урожайности ячменя за 1982 год, не использованные ранее, который был выращен в регионе Меньково. Аналогично предыдущему эксперименту, все необходимые данные, такие как погодные условия, вегетативные показатели и т.п. были получены с метеостанции Белогорка Ленинградской области.

Таблица, с идентичными параметрами, имеет размерность  $349 \times 32$  (строки  $\times$  столбцы). (см. Рисунок 2.1.4)

	Sbiom#1	Ebiom#1	SumBiom#1	Rbiom#1	PhTime#1	LAI#1	PrimAss#1	ave#1	SumRad#1	Prec#2	...	SumNDen#1	SumNLow#1	SumNPlant#1	WPotSe
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	13.50	367.934	0.0	...	0.0000	0.0010	0.0000	247.
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	15.30	294.835	0.0	...	0.0000	0.0019	0.0000	261.
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	14.80	288.245	7.4	...	0.0000	0.0028	0.0000	249.
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	16.25	296.330	2.7	...	0.0000	0.0037	0.0000	222.
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	15.65	270.744	0.7	...	0.0000	0.0045	0.0000	217.
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
344	54.593	54.686	144.497	-9.970	1.927	0.514	2.504	19.50	537.211	0.0	...	0.4665	11.5963	169.3533	9318.
345	54.322	54.942	144.003	-9.902	1.947	0.371	1.423	20.45	432.806	0.0	...	0.4665	11.5958	169.4570	9312.
346	53.980	54.952	143.212	-9.836	1.971	0.201	1.085	20.00	469.701	0.0	...	0.4665	11.5953	168.8699	9469.
347	53.539	54.634	142.019	-9.774	2.000	0.002	0.398	19.40	511.151	0.0	...	0.4665	11.5948	167.9443	9676.
348	53.151	54.238	140.909	-9.727	2.027	-0.184	0.000	15.50	355.686	5.9	...	0.4665	11.5943	166.5571	9057.

349 rows × 32 columns

Рисунок 2.1.4 - Таблица, использованная для второго численного эксперимента.

Обучение модели на новых данных было выполнено по предыдущей схеме. Для выбора параметра регуляризации использовалась перекрестная проверка GridSearchCV. Выбор проходил между переменными [0, 0.0001, 0.001, 0.01, 0.1, 0.5, 1]. Модель определила  $\alpha=0.001$  как наилучший параметр регуляризации, который является аналогичным предыдущему результату эксперимента перекрестной проверки. Оценщики модели MAE, RMSE,  $R^2$  при выбранном параметре регуляризации показали данный результат (Таблица 2):

Таблица 2. Результат модели на новых данных с параметром регуляризации, выбранным в результате перекрестной проверки.

Оценщик моделей	Результаты	
$R^2$ (коэффициент детерминации)	1.000 ( <i>Train</i> )	0.952 ( <i>Test</i> )
RMSE (среднеквадратическое отклонение)	0.21937518202085762	
MAE (средняя абсолютная ошибка)	0.19417394299829732	

Результаты на новых данных, представленные в приведенной выше таблице, показывают, что модель регрессии LASSO и при использовании ранее неизвестных данных, объем которых превышает предыдущие в 3 раза, имеет высокие показатели точности. При этом модель обучалась лишь на 15% данных от общего.

Успешность прогноза можно определить, основываясь на показателях коэффициента детерминации  $R^2$ , который на тренировочных данных показал результат равный 1, а на тестовых равный 0.952. Идеальный результат прогноза на тренировочных данных можно объяснить тем, что модель ранее имела опыт обучения. К тому же, количество тренировочных данных было уменьшено с 50 процентов до 15, то есть почти в три с половиной раза. На тестовых данных

модель показала результат точности более 95 процентов, что также определяет успешность прогноза регрессии LASSO.

MAE (средняя абсолютная ошибка) между прогнозируемыми значениями и фактическими значениями для ячменя составляет 0.194. А RMSE (среднеквадратическое отклонение) между прогнозируемой и фактической урожайностью составила 0.219. Основываясь на результатах двух выполненных численных экспериментах, очевидно, что регрессия LASSO является подходящим инструментом для прогнозирования урожайности ячменя.

Ниже представлена визуализация результатов второго численного эксперимента, полученная с помощью библиотеки matplotlib. (Рисунок 2.1.5)

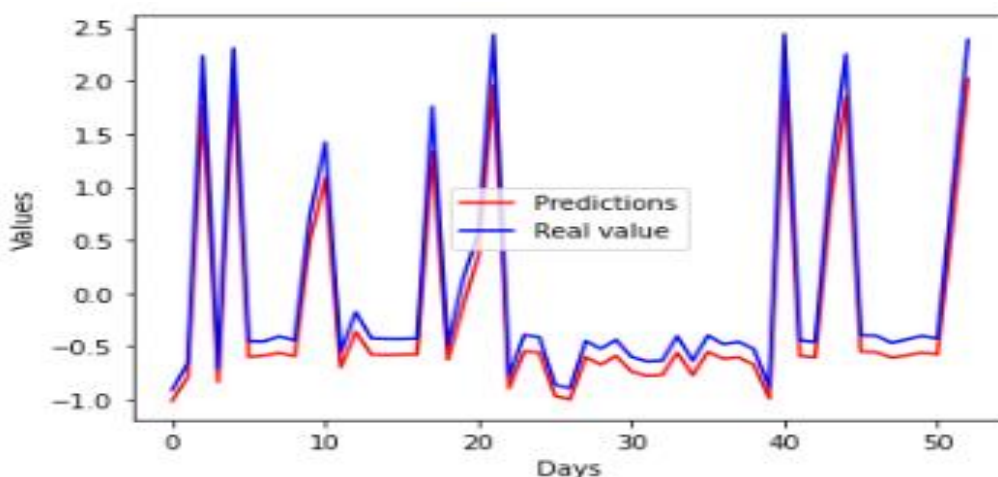


Рисунок 2.1.5 - Визуализация спрогнозированных и реальных значений урожайности второго численного эксперимента (Красная линия - спрогнозированные значения, синяя линия – реальные значения)

## 2.2 Важность переменных

Корреляция между урожайностью и погодными параметрами была посчитана в програмном комплексе “ AGROTOOL v.3.5”, посредством графической визуализации результатов расчета в полностью настраиваемой и управляемой в режиме RunTime среде, как было описано ранее в разделе 1.7.

Среднее количество осадков, выпавшее за весь период, с мая по август 1988 года, взращивания ячменя, составляет 250 мм. Количество осадков варьируются в диапазоне 0–650 мм.

Ниже показан график осадков за весь период (Рисунок 2.2.1) и их корреляция с урожайностью (Рисунок 2.2.2).

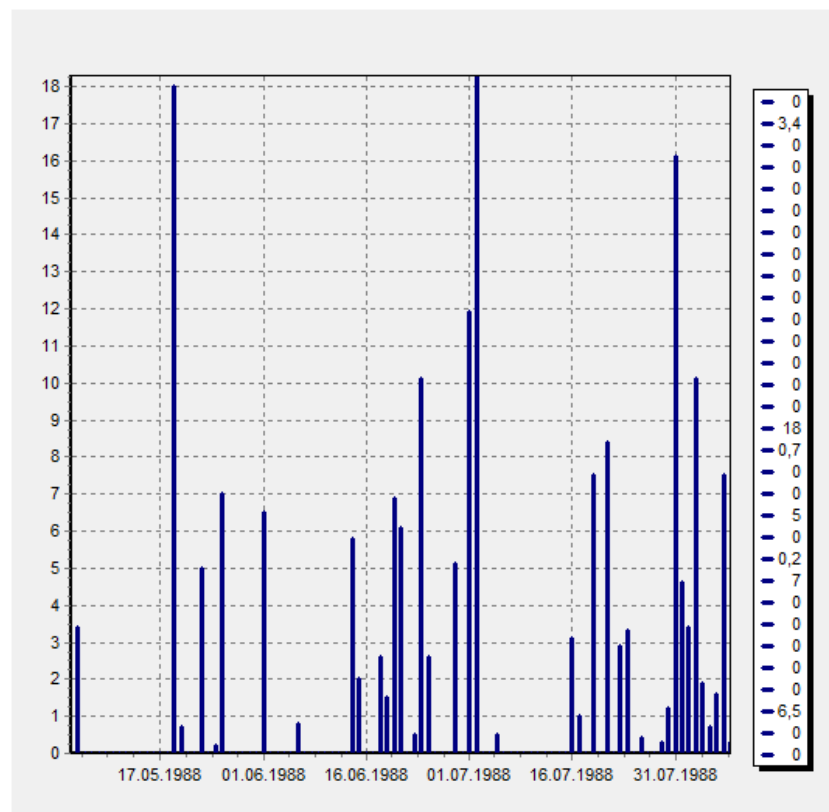


Рисунок 2.2.1 - График осадков в периоде с мая по август 1988 года.

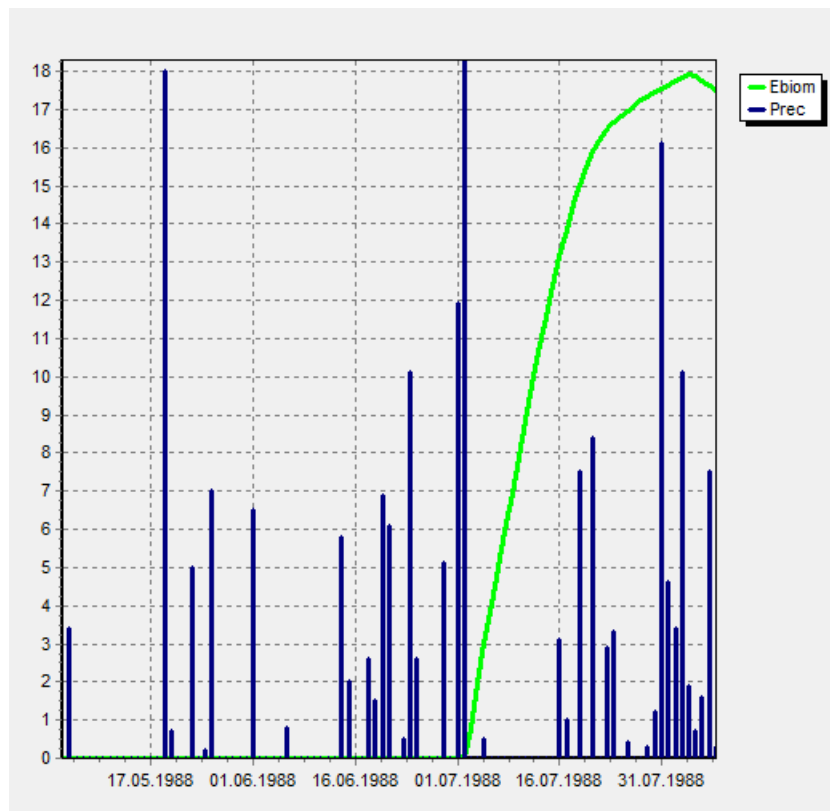


Рисунок 2.2.2 - Корреляция осадков и урожайности.



Как видно на Рисунке 2.2.2, осадки вносят существенный вклад в урожай ячменя в исследуемом регионе. В течение первого месяца после посадки изучаемой культуре не требуется много осадков по сравнению с последующими месяцами. Но, тем не менее, осадки в первые месяцы посадки влияют на скорость первоначального прорастания и формирование вегетативных органов. На графике видно, что период с начала посадки до 01.07.1988 имеет более обильные осадки, по сравнению с периодом от 01.07.1988 до сбора урожая. Очевидно, что с 1.07.1988 развитие роста генеративного органа происходит стремительно быстрее. Из чего можно сделать вывод, что осадки в первые месяцы, хоть и не требуются в большом количестве, но имеют влияния на прорастание в дальнейших фазах.

Аналогичным образом, анализ требований к температуре растений был проведен с использованием графической среды “AGROTOOL v.3.5”.

Как было описано ранее в разделе 1.5, оптимальной температурой для выращивания ячменя и получения хорошего урожая варьируется от фазы прорастания. Во время созревания она составляет 20-23 °С. Максимальная требуемая температура должна быть в пределах 1–12 °С в течение первого месяца, в пределах 12–20 °С в течение 2-го месяца и в диапазоне 20–23 °С в течение 3-го месяца полного созревания ячменя. Полученный график показывает, как варьируется температура в течение всего периода прорастания. Судя по графику, можно сделать вывод, что температура в данной местности является приемлемой для получения оптимального урожая ячменя.

График изменений температуры за весь период выращивания ячменя показан ниже на рисунке 2.2.3.

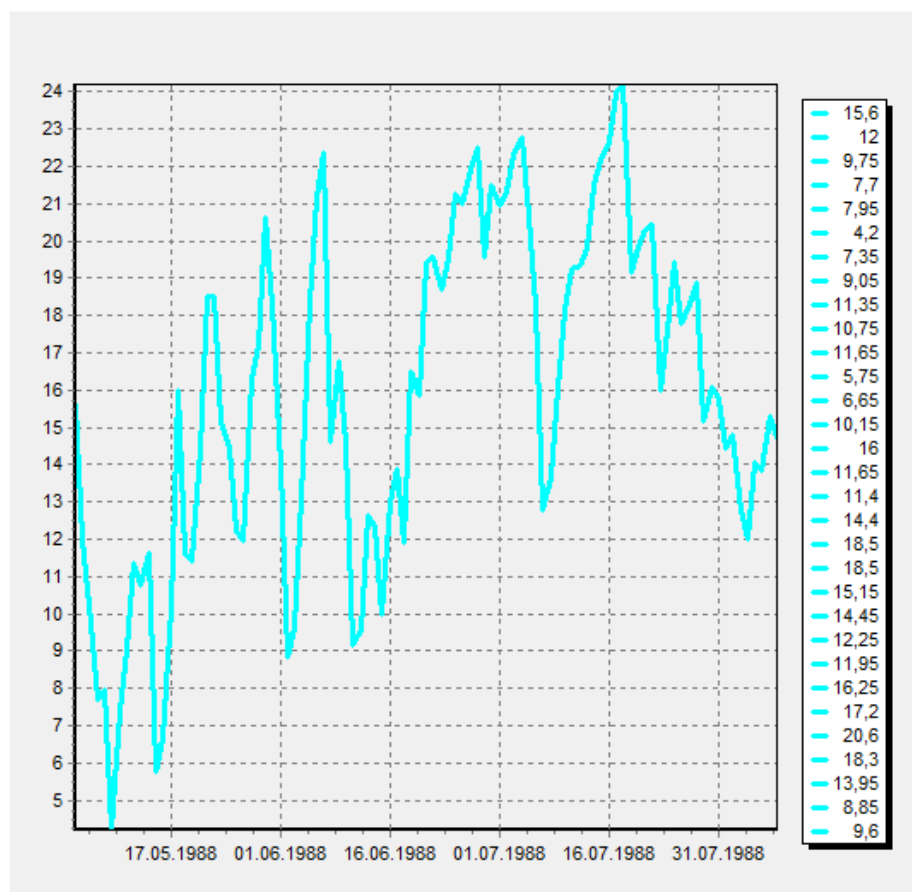


Рисунок 2.2.3 - Показатели значений температуры на период взращивания ячменя.

### 2.3 Выводы результатов

Урожайность сельскохозяйственных культур зависит как от контролируемых, так и от неконтролируемых факторов. К первым относятся такие факторы, как сорта культур или семян, методы обработки почвы, использование удобрений и многие другие. К неконтролируемым относятся факторы, не зависящие от человека. Например, погодные переменные, то есть осадки, температура воздуха, солнечная радиация, температура почвы, влажность воздуха, влажность почвы, и т. д. Если какой-либо из параметров превышает или ниже уровня требований растения, то оно может плохо расти, и это повлияет на продуктивность.

Для получения хорошего прогноза урожайности очень важно использовать вышеупомянутые погодные переменные. Однако, это может быть ограничено доступностью данных. По этой причине в данной работе данные об осадках и температуре использовались за 1988 год для прогнозирования урожайности ячменя в качестве синтетических данных для того, чтобы определить эффективность используемого метода – регрессии LASSO.

Требования к количеству осадков и температуре различаются от одной стадии выращивания к другой. Это означает, что неадекватная погодная переменная на

данном этапе повлияет на уровень производства. В текущем исследовании ежемесячное совокупное количество осадков и значения средней дневной температуры использовались для отражения потребностей ячменя на каждой стадии выращивания урожая

## **2.4 Стадии выращивания сельскохозяйственных культур и климатические требования**

Количество воды, необходимое любому растению, зависит от стадии выращивания культуры, потребности атмосферы в испарении, вида культуры и других параметров.

Ячмень проходит различные стадии со дня посадки: прорастание, восходы, кущение, выход в трубку, колошение, цветение и созревание (от молочной спелости до полного созревания)

Урожайность ячменя составила 17,502 ц/га. Стадия прорастания занимает 26 дней, что означает, что это происходит в течение первого месяца и для получения оптимального урожая количество осадков было в диапазоне 35–180 мм, а температура была 4–18,5°. С последней недели 1-го месяца до конца 2-го месяца проходит стадия вегетативного роста и колошение (фаза развития колоса). Согласно анализу, осадки и температура для этих этапов составляют 26–102 мм и 18–21 °С соответственно. На стадии цветения оптимальные осадки и температура для урожая составили 50–185 мм и 13–23 °С соответственно, тогда как на стадии созревания осадки и температура для данного урожая составили 10–250 мм и 14–24 °С соответственно.

## **2.5 Прогнозирование и производительность моделей**

Прогнозирование урожайности ячменя с помощью машинного обучения является важным экспериментом, поскольку он помогает предоставить информацию о тенденциях урожайности и может быть полезен для принятия решений.

В этом исследовании набор данных по ячменю был обучен моделью регрессии LASSO с целью определения её эффективности в качестве предсказателя урожайности. В качестве основных предикторов выступали количество осадков и температура, по которым делались основные выводы требований культуры к погодным переменным. Но такие переменные как влажность воздуха, влажность почвы, солнечная радиация и т. д. учитывались, так как имеют большое влияние на урожайность культуры.

Результаты обученной и протестированной моделей показывают, что точность для прогноза урожайности ячменя с помощью регрессии LASSO составил 17,3 ц/г. Дисперсия между прогнозируемой и фактической

урожаемостью оценивалась через  $R^2$  (коэффициент детерминации) и была равна 0,987.

Результаты, полученные в ходе двух численных экспериментов по прогнозированию урожайности ячменя, показывают, что модель следует рекомендовать при разработке системы раннего прогнозирования урожайности. Пробное исследование, в котором использовались новые данные большего количества, также показала эффективность прогноза регрессионной модели. Во втором пробном численном эксперименте рассматривались климатические переменные, так как влияние изменения климата на урожайность сельскохозяйственных культур имеет решающее значение. На различных фазах прорастания сельскохозяйственные культуры могут подвергаться влиянию изменения климата, что, впоследствии, может привести к снижению урожайности. По этой причине мы рассматривали количество осадков и температуру, как основные переменные, также учитывая и другие переменные, как влажность воздуха, почвы, количество радиации, длина дня и т. п. Опираясь на показатели основных переменных, которыми были определены осадки и температура, были выведены графики их корреляции на каждой стадии роста ячменя.

Несмотря на то, что регрессия LASSO показала свою эффективность при прогнозировании урожайности ячменя, это исследование было ограничено использованием только одной модели. Улучшение результатов прогнозирования можно добиться, используя несколько моделей для прогноза и определить оптимальный для каждого типа экспериментов.

## ЗАКЛЮЧЕНИЕ

В данном эксперименте изучалось влияние изменения климата на производство сельскохозяйственной культуры - ячменя. Работа выступает за ранний обмен информацией, в частности, об ожидаемом урожае, чтобы обеспечить надлежащее планирование, которое может помочь уменьшить отсутствие продовольственной безопасности.

Регрессия LASSO была протестирована на данных из Меньково (сельскохозяйственная деревня Ленинградской области) на предмет её способности прогнозировать урожайность ячменя. Метод регрессии LASSO показал свою эффективность, о чем свидетельствуют ее значения коэффициента детерминации ( $R^2=0.987$ ), среднеквадратического отклонения ( $RMSE=0.113$ ) и средней абсолютной ошибки ( $MAE=0.081$ ).

Исторические данные об урожайности сельскохозяйственной культуры выступали в качестве целевой переменной, а погодные значения использовались в качестве предикторов. Кроме того, проведен анализ корреляции между погодными переменными и урожайностью сельскохозяйственной культуры.

В дальнейшем разнообразные методы машинного обучения также можно использовать для получения лучших результатов прогнозирования не только урожайности, но и событий других отраслей, необходимых для обеспечения безопасности жизнедеятельности человека.

Методы машинного обучения доказали свою ценность в качестве решения задач прогнозирования благодаря своей способности извлекать значимую информацию из измеримых данных. И могут служить отличным помощником для облегчения задач прогнозирования и планирования.

## СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

- 1 Бейлюэн, Д.; Шаубергер, Б.; Бастос, А.; Сиаис, П.; Маковски, Д. Влияние экстремальных погодных условий на производство сельскохозяйственных культур в Европе в 2018 году: случайный лес — аномалии урожайности. Филос. Транс. Р. Соц. Б биол. науч. 2020. - 375с.
- 2 Джавадинежад, С. ; Эсламян, С.; Аскари, К.А. Анализ важнейших климатических параметров, влияющих на показатели изменчивости урожая в условиях меняющегося климата. Междунар. Дж. Гидрол. науч. Технол. 2021.- 1–25с.
- 3 Кин, Б.А. Погода и урожай. QJR Метеорол. соц. 1940/ - 155–166 с.
- 4 Крейн-Дреш, А. Методы машинного обучения для прогнозирования урожайности и оценки воздействия изменения климата в сельском хозяйстве. Окружающая среда. Рез. лат. 2018 - 13с.
- 5 Ку, С.-Б.; Эдвардс, GE; Таннер, К.Б. Влияние света, углекислого газа и температуры на фотосинтез, кислородное ингибирование фотосинтеза и транспирацию у *Solanum tuberosum*. Завод Физиол. 1977.
- 6 Прогнозирование урожайности сельскохозяйственных культур: рабочая тетрадь / Г. Ф. Петрик, А. Г. Прудников, Т. В. Логойда – Краснодар: КубГАУ, 2017 – 49 с.
- 7 Прогнозирование урожайности сельскохозяйственных культур: учеб. Пособие А. И. Трубилин, Г. Ф. Петрик, А. Г. Прудников – Краснодар : КубГАУ, 2017 – 95 с.
- 8 Шакур, У.; Сабур, А.; Али, И.; Мохсин, А.К. Влияние изменения климата на сельское хозяйство: эмпирические данные из засушливого региона, Пакистан. Дж. Агр. науч. 2011. - 327–333. [Академия Google]
- 9 Ю, С.; Лим, Х.; Хео, Дж. Подходы к машинному обучению для прогнозирования урожайности с помощью MODIS и данных о погоде. В материалах 40-й Азиатской конференции по дистанционному зондированию: прогресс технологии дистанционного зондирования для «умного будущего», ACRS 2019, Тэджон, Республика Корея, 14–18 октября 2019. - 1–4с.
- 10 Общие географические данные Ленинградской области // Сайт // <https://potatosystem.ru/kartofelevodstvo-leningradskoj-oblasti/> .
- 11 Документация программного комплекса “AGROTOOL” // <http://www.rpoluektov.ru/%D0%BC%D0%BE%D0%B4%D0%B5%D0%BB%D1%8C-agrotool/> .
- 12 Географическая карта Ленинградской области // [https://fedoroff.net/load/maps/karta/geograficheskaja\\_karta\\_leningradskoj\\_oblasti/90-1-0-1540](https://fedoroff.net/load/maps/karta/geograficheskaja_karta_leningradskoj_oblasti/90-1-0-1540)
- 13 География Ленинградской области // Геологическое строение и полезные ископаемые // [https://ru.wikipedia.org/wiki/География\\_Ленинградской\\_области#Геологическое\\_строение\\_и\\_полезные\\_ископаемые](https://ru.wikipedia.org/wiki/География_Ленинградской_области#Геологическое_строение_и_полезные_ископаемые).

14 Сельское хозяйство Ленинградской области // <https://abcentre.ru/page/selskoe-hozyaystvo-leningradskoy-oblasti> .

15 Российская академия сельскохозяйственных наук // [https://ru.wikipedia.org/wiki/Российская\\_академия\\_сельскохозяйственных\\_наук](https://ru.wikipedia.org/wiki/Российская_академия_сельскохозяйственных_наук) .

16 Регрессия LASSO и Ridge // [https://ru.wikipedia.org/wiki/%D0%9C%D0%B5%D0%BD%D1%8C%D0%0%BE%D0%B2%D0%BE\\_\(%D0%9B%D0%B5%D0%BD%D0%B8%D0%BD%D0%B3%D1%80%D0%B0%D0%B4%D1%81%D0%BA%D0%B0%D1%8F\\_%D0%BE%D0%B1%D0%BB%D0%B0%D1%81%D1%82%D1%8C\)](https://ru.wikipedia.org/wiki/%D0%9C%D0%B5%D0%BD%D1%8C%D0%0%BE%D0%B2%D0%BE_(%D0%9B%D0%B5%D0%BD%D0%B8%D0%BD%D0%B3%D1%80%D0%B0%D0%B4%D1%81%D0%BA%D0%B0%D1%8F_%D0%BE%D0%B1%D0%BB%D0%B0%D1%81%D1%82%D1%8C)) .

17 5 алгоритмов регрессии // Интернет-источник // <https://habr.com/ru/companies/vk/articles/513842/> .

18 Почему Python лучший в машинном обучении // Интернет-статья// <https://pythonist.ru/pochemu-python-luchshij-yazyk-dlya-mashinnogo-obucheniya-i-ii/>.

## Приложение А

```
1 import matplotlib.pyplot as plt # библиотека для визуализации
2 import numpy as np # Для работы с многомерными массивами
3 import pandas as pd # Для обработки и анализа данных
4 from pandas import DataFrame as df
5 from sklearn.model_selection import train_test_split # Для разделения датасета перед обучением
6 from sklearn.preprocessing import StandardScaler # Стандартизация данных
7

1 path=r'C:\Users\Ailee\agrootool\DataBases\ODB\ODB.xls'
2 dataset=pd.read_excel(path)
3 print(dataset.shape)

(7, 9)

1 df = pd.read_excel(r'C:\Users\Ailee\agrootool\DataBases\ODB\ODB.xls',
2 sheet_name='Results')

1 df=df.drop(columns=['RESULTS'],axis=1)

1 df=df.dropna()

1 df.columns=df.iloc[0]
2 df=df.drop(index=0,axis=0)

1 data=df.copy(deep=True)

1 data['DATE'] = pd.to_datetime(data['DATE']).astype('int64')

1 data=data.drop(columns=['IPhase#4'],axis=1)
2

1 data.to_excel(r'C:\Users\Ailee\final_data.xlsx')
```

Рисунок А.1 – Предобработка данных, перезапись в новый файл «final data»

## Продолжение А



```

1 path2=r'C:\Users\Ailee\agrotool\Final_data.xlsx'
2 final_data=pd.read_excel(path2)
3 print(final_data.shape)

```

(97, 38)

```

1 x=final_data.iloc[:, 3:38].values
2 y=final_data.iloc[:,2].values

```

```
1 final_data.iloc[:, 3:38]
```

	Lbiom#1	Sbiom#1	SumBiom#1	Rbiom#1	PhTime#1	LAI#1	PrimAss#1	ave#1 <sup>T</sup>	SumRad#1	Prec#2	...	SumNLow#1	SumNPlant#1	ExtNPool#1	WPo
0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	15.60	487.677433	0.0	...	0.005019	0.000000	0	125
1	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	12.00	274.940949	3.4	...	0.009409	0.000000	0	127
2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	9.75	428.752550	0.0	...	0.013507	0.000000	0	130
3	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	7.70	514.201380	0.0	...	0.017374	0.000000	0	141
4	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	7.95	438.340830	0.0	...	0.021058	0.000000	0	154
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
92	4.692697	17.516174	40.118497	-6.053531	1.935817	0.053926	0.011177	12.00	307.582951	1.9	...	0.035187	48.876434	0	587
93	4.651703	17.427000	39.943634	-6.027090	1.955582	0.037153	0.000000	14.05	266.526527	0.7	...	0.034641	48.876434	0	561
94	4.611626	17.314394	39.675515	-6.001126	1.975150	0.020689	0.000000	13.85	276.199210	1.6	...	0.034121	48.665923	0	561
95	4.567694	17.190686	39.381059	-5.972542	1.995602	0.003644	0.000000	15.30	298.202698	7.5	...	0.033628	48.348528	0	513
96	4.525953	17.072866	39.100717	-5.945252	2.015730	-0.012972	0.000000	14.70	294.803530	0.3	...	0.033162	47.999913	0	478

97 rows × 34 columns

```
1 final_data.iloc[:,2]
```

```

0    0.000000
1    0.000000
2    0.000000
3    0.000000
4    0.000000
...
92    17.909627
93    17.864931
94    17.749495
95    17.622679
96    17.501897

```

Name: Ebiom#1, Length: 97, dtype: float64

```
1 x_train, x_test, y_train, y_test=train_test_split(x,y, test_size=0.5, random_state=0)
```

```

1 sc_data=StandardScaler()
2 x_train=sc_data.fit_transform(x_train)
3 x_test=sc_data.transform(x_test)
4 #pd.DataFrame(x_train)
5 #pd.DataFrame(x_test)

```

```

1 y_train_sc_data=np.array(y_train).reshape(-1,1)
2 y_train_sc_data=sc_data.fit_transform(y_train_sc_data)
3 #pd.DataFrame(y_train_sc_data)

```

```

1 y_test_sc_data=np.array(y_test).reshape(-1,1)
2 y_test_sc_data=sc_data.fit_transform(y_test_sc_data)
3 #pd.DataFrame(y_test_sc_data)

```

Продолжение А

```

1 from sklearn.model_selection import GridSearchCV, KFold, cross_val_score
2 folds=KFold(n_splits=5, shuffle=True, random_state=42)
3 params={'alpha': [0.001, 0.01, 0.1, 0.5, 1]}
4 model=Lasso()
5 model_cv= GridSearchCV(estimator=model, param_grid=params, scoring='r2', cv=folds, return_train_score=True)
6 model_cv.fit(x_train, y_train_sc_data)

GridSearchCV(cv=KFold(n_splits=5, random_state=42, shuffle=True),
  estimator=Lasso(),
  param_grid={'alpha': [0.001, 0.01, 0.1, 0.5, 1]},
  return_train_score=True, scoring='r2')

1 model_cv.best_params_

{'alpha': 0.001}

```

Рисунок А.2 - Разделение данных на тестовые и тренировочные и их масштабирование.

```

1 lasso= Lasso(alpha=0.001)
2 lasso.fit(x_train, y_train_sc_data)
3 y_train_pred_data_lasso = lasso.predict(x_train)
4 y_test_pred_data_lasso = lasso.predict(x_test)
5 print(lasso.coef_)
6
7 print('R^2 train: {:.3f}, test: {:.3f}'.format(
8     metrics.r2_score(y_train_sc_data, y_train_pred_data_lasso),
9     metrics.r2_score(y_test_sc_data, y_test_pred_data_lasso)))
10 print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(y_test_sc_data, y_test_pred_data_lasso)))
11 print('Mean Absolute Error:', metrics.mean_absolute_error(y_test_sc_data, y_test_pred_data_lasso))
12
[-0.00000000e+00 -2.43163538e-01  5.12331894e-01  0.00000000e+00
 3.83040995e-01 -2.53585078e-01 -0.00000000e+00  2.35574922e-03
-2.75874460e-02 -1.08981716e-02 -0.00000000e+00 -0.00000000e+00
-1.11446284e-04  1.45475610e-02 -4.52193657e-02  0.00000000e+00
 0.00000000e+00  1.83909167e-01  1.46452490e-01 -3.15180875e-02
-0.00000000e+00 -0.00000000e+00  2.22209247e-03  0.00000000e+00
-0.00000000e+00  3.65502580e-02  0.00000000e+00  1.80255655e-03
 0.00000000e+00  0.00000000e+00  1.30188264e-02 -0.00000000e+00
-6.15306353e-02  0.00000000e+00]
R^2 train: 0.999, test: 0.987
Root Mean Squared Error: 0.11338038219738471
Mean Absolute Error: 0.07974589008840262

1 plt.plot(y_test_pred_data_lasso, "r")
2 plt.plot(y_test_sc_data, "b")
3 plt.legend(["Predictions", "Real value"])
4 plt.xlabel("Days")
5 plt.ylabel("Values")
6 plt.show()

```

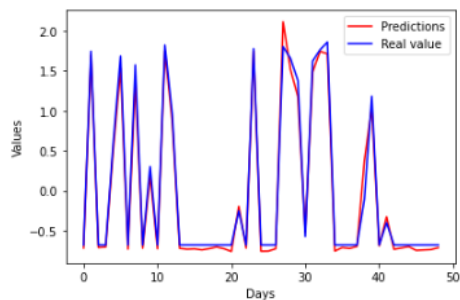


Рисунок А.3 – Реализация регрессии LASSO и визуализация результата.