



Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное
Образовательное учреждение высшего образования
«Национальный исследовательский Томский политехнический университет (ТПУ)»

Школа Инженерная Школа Ядерных Технологий

Направление подготовки 01.03.02 Прикладная Математика и Информатика

ООП/ОПОП _____

Отделение школы (НОЦ) Отделение Экспериментальной Физики

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА БАКАЛАВРА

Тема работы
<i>Проведение многомерного регрессионного анализа с целью предсказания урожайности картофеля</i> УДК 519.237.5:635.21.075

Обучающийся

Группа	ФИО	Подпись	Дата
0В91	Чернюгов Кирилл Викторович		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Семенов М. Е.	к.ф.-м.н., доцент		

Со-Руководитель ВКР (по разделу «Концепция стартап проекта»)

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор	Чистякова Н.О.	д.э.н		

КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Антоневич О.А.	д.б.н		

ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП/ОПОП, должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Крицкий О.Л	к.ф.-м.н., доцент		

Министерство науки и высшего образования Российской Федерации
федеральное государственное автономное образовательное учреждение
высшего образования
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Инженерная школа ядерных технологий
Направление подготовки 01.03.02 «Прикладная математика и информатика»
Отделение экспериментальной физики

УТВЕРЖДАЮ:
Руководитель ООП
_____ Крицкий О.Л.
(Подпись) (Дата) (Ф.И.О.)

ЗАДАНИЕ
на выполнение выпускной квалификационной работы

В форме:

Бакалаврской работы

Студенту:

Группа	ФИО
0В92	Чернюгову Кириллу Викторовичу

Тема работы:

Проведение многомерного регрессионного анализа с целью предсказания урожайности картофеля	
Утверждена приказом директора (дата, номер)	

Срок сдачи студентом выполненной работы: 15.06.2023

ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

Исходные данные к работе	<i>Данные ежегодных замеров химического состава почвы и урожайности близости города Квебек</i>
Перечень подлежащих исследованию, проектированию и разработке вопросов	<ol style="list-style-type: none">1. Провести анализ датасета2. Провести исследование климатических зон РФ, на возможность соответствия датасету3. Провести анализ нескольких регрессионных алгоритмов, с целью выяснения наиболее оптимального для реализации проекта4. Обучение модели5. Реализация метода на выбранном языке программирования.6. Проведение сравнения предсказанных характеристик с реальными.
Перечень графического материала	<ol style="list-style-type: none">1. Визуализация выбранной модели2. Диаграммы методов обучения.

	3. Скриншоты работы программного продукта. 4. Представления работы метода в математическом виде
Консультанты по разделам выпускной квалификационной работы <i>(если необходимо, с указанием разделов)</i>	
Раздел	Консультант
Концепция стартапа	Чистякова Н.О.

Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику	15.03.2023
---	------------

Задание выдал руководитель:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОЭФ	Семенов Михаил Евгеньевич	к. ф.-м. н., доцент		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
0В91	Чернюгов Кирилл Викторович		

**ЗАДАНИЕ К РАЗДЕЛУ
«КОНЦЕПЦИЯ СТАРТАП-ПРОЕКТА»**

Обучающемуся:

Группа	ФИО
0B91	Чернюгов Кирилл Викторович

Школа	ИЯТШ	Отделение школы (НОЦ)	Отделение Экспериментальной Физики
Уровень образования	Бакалавриат	Направление/ООП/ОПОП	01.03.02«Прикладная математика и информатика»

Перечень вопросов, подлежащих разработке:	
<i>Проблема конечного потребителя, которую решает продукт, который создается в результате выполнения НИОКР (функциональное назначение, основные потребительские качества)</i>	Решение проблемы невозможности автоматического мониторинга химического состава почвы с целью отслеживания на урожайность
<i>Способы защиты интеллектуальной собственности</i>	Товарный знак Патент вида «полезная модель»
<i>Объем и емкость рынка</i>	5,3 млрд рублей
<i>Современное состояние и перспективы отрасли, к которой принадлежит представленный в ВКР продукт</i>	Рынок проявляет сильные тенденции к цифровизации, а из-за набора исключительно иностранных решений, появляется потребность в импортозамещении
<i>Себестоимость продукта</i>	1283,11 руб/час
<i>Конкурентные преимущества создаваемого продукта и Сравнение технико-экономических характеристик продукта с отечественными и мировыми аналогами</i>	Цена Возможности
<i>Целевые сегменты потребителей создаваемого продукта</i>	Сельскохозяйственный центр Крупный аграрный сектор
<i>Бизнес-модель проекта, производственный план и план продаж</i>	Представлен в виде таблицы Остервальдера
<i>Стратегия продвижения продукта на рынок</i>	Гос. Контракты Выставки Развитие бренда
Перечень графического материала:	
1. ...	1) Таблицы
2. ...	2) Графики

Дата выдачи задания к разделу в соответствии с календарным учебным графиком	29.04.23
--	----------

Задание выдал консультант по разделу «Концепция стартап-проекта» (со-руководитель ВКР):

Должность	ФИО	Ученая степень, звание	Подпись	Дата
------------------	------------	-------------------------------	----------------	-------------

Профессор	Чистякова Н.О.	д.э.н		
-----------	----------------	-------	--	--

Задание принял к исполнению обучающийся:

Группа	ФИО	Подпись	Дата
0В91	Чернюгов Кирилл Викторович		

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа		ФИО	
0В91		Чернюгов Кирилл Викторович	
Школа	ИЯТШ	Отделение (НОЦ)	Экспериментальной физики
Уровень образования	Бакалавриат	Направление/специальность	01.03.02 Прикладная Математика и Информатика

Тема ВКР:

Проведение многомерного регрессионного анализа для предсказания урожайности картофеля

Исходные данные к разделу «Социальная ответственность»:

Введение

- Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика) и области его применения.
- Описание рабочей зоны (рабочего места) при разработке проектного решения/при эксплуатации

Объект исследования Датасет с данными о составе химической почвы и урожайности
Область применения Сельское хозяйство
Рабочая зона: офис
Размеры помещения: 20*30 м

Количество и наименование оборудования рабочей зоны
 Офис: Ноутбук(1 шт.), Сервер (1 ед.)
Рабочие процессы, связанные с объектом исследования, осуществляющиеся в рабочей зоне
 Офис: Разработка ПО для обработки результата химического анализа почвы

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. Правовые и организационные вопросы обеспечения безопасности при разработке проектного решения

- специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства;
- организационные мероприятия при компоновке рабочей зоны.

Правовые нормы трудового законодательства:
 СанПиН 2.2.4.3359-16 "Санитарно-эпидемиологические требования к физическим факторам на рабочих местах" (утв. постановлением Главного государственного санитарного врача РФ от 21 июня 2016 г. № 81)
 Типовая инструкция по охране труда при работе на персональном компьютере (ТОИ Р-45-084-01, утв. Приказом Минсвязи РФ от 02.07.2001 N 162)
Организационные мероприятия при компоновке рабочей зоны:

ГОСТ 12.2.032-78 ССБТ «Рабочее место при выполнении работ сидя. Общие эргономические требования».

СанПиН 1.2.3685-21 «Гигиенические требования к персональным электронно-вычислительным машинам и организации работы» - рабочие места с ПК по отношению к световым проемам должны располагаться так, чтобы естественный свет падал сбоку, желательно слева.

ГОСТ 12.2.032-78. «Система стандартов безопасности труда». Рабочее место при выполнении работ сидя» рабочий стол может быть любой конструкции, отвечающей современным требованиям эргономики и позволяющей удобно разместить на рабочей поверхности оборудование с учетом его количества, размеров и характера выполняемой работы.

<p>2. Производственная безопасность при разработке проектного решения):</p> <ul style="list-style-type: none"> – Анализ выявленных вредных и опасных производственных факторов 	<p><u>Вредные и опасные факторы (ГОСТ 12.0.003-2015):</u></p> <ul style="list-style-type: none"> • Отклонение показателей микроклимата в закрытом помещении • Недостаточная освещенность • Нервно-психические перегрузки, монотонность трудового процесса • Повышенное значение напряжения в электрической цепи, замыкание которой может произойти через тело человека
<p>3. Экологическая безопасность при разработке проектного решения/при эксплуатации</p>	<p>При написании ВКР вредных выбросов в атмосферу, почву и водные источники не производилось, радиационного заражения не произошло, чрезвычайные ситуации не наблюдались, поэтому не оказывались существенные воздействия на окружающую среду, и никакого ущерба окружающей среде не было нанесено.</p> <ul style="list-style-type: none"> • <u>Утилизация бумажных документов:</u> <ol style="list-style-type: none"> 1. Сжигание документов. 2. Шредирование. 3. Закапывание. 4. Химическая обработка. • <u>Утилизация ПЭВМ:</u> <p>ГОСТ Р53692-2009 Ресурсосбережение</p> <p>Вышедшее из строя ПЭВМ и сопутствующая оргтехника относится к IV классу опасности и подлежит специальной утилизации. Для оказания наименьшего влияния на окружающую среду, необходимо проводить специальную процедуру утилизации ПЭВМ и оргтехники, при которой более 90% отправится на вторичную переработку и менее 10% будут отправлены на свалки.</p> • Ртуть из разбитых люминесцентных ламп под влиянием бактерий на полигонах ТБО преобразуется в токсичную водорастворимую метилртуть. Она накапливается в грунтах, открытых водоемах и подземных источниках, а затем – в организмах рыб и моллюсков. <p>Метилртуть, которую человек получает из пищи рыбного происхождения, отравляет нервную систему, наносит непоправимый вред пищеварению и иммунитету, легким и почкам. Если ее концентрация зашкалит, возможен и летальный исход. При умеренной интоксикации наблюдаются постоянные головные боли, потеря памяти, бессонница, нарушения моторной и когнитивной функции.</p> <p>Отсюда возможно сильное влияние на экологию в зоне гидросферы</p> <ul style="list-style-type: none"> • В ходе деятельности организация также создает бытовой мусор (канцелярские,

	<p>пищевые отходы, искусственные источники освещения), который должен быть утилизирован в соответствии с определенным классом опасности или переработан, чтобы не оказывать негативное влияние на состояние литосферы.</p>
<p>4. Безопасность в чрезвычайных ситуациях при разработке проектного решения</p>	<p>Возможные ЧС:</p> <ul style="list-style-type: none"> • Природные катастрофы (наводнения, цунами, ураган и т.д.); • Геологические воздействия (землетрясения, оползни, обвалы, провалы территории и т.д.) • Техногенные аварии (Пожар) <p>Наиболее типичная ЧС: Пожар</p>
<p>Дата выдачи задания для раздела по линейному графику</p>	

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Антоневич Ольга Алексеевна	к.б.н		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
0B91	Чернюгов Кирилл Викторович		

РЕФЕРАТ

Выпускная квалификационная работа содержит 73 листов, 28 рисунков, 49 источников, 10 таблиц.

Ключевые слова: Регрессионный анализ, Картофель, Химический состав, Урожайность, Нейронные сети.

Объектом исследования является датасет содержащий данные о характеристиках вида картофеля, химического состава почвы, а также урожайности

Цель работы – провести регрессионный анализ данных с целью формирования модели оценки предсказания урожайности картофеля.

Методы исследования: Регрессионный анализ – Модель Random Forest, модель Linear Regression и Модель Multi-Layer Perceptron, также использовался коэффициентный анализ оценок моделей.

В результате исследования проведен анализ химического состава почвы и характеристик картофеля, на основе использования данных, проведена комплексная оценка эффективности моделей, разработана программа, реализующая данные о составе почвы и характеристиках овоща, а также результаты обучения регрессионной модели в оценки предсказания урожайности. Дана интерпретация полученных результатов и сделаны соответствующие выводы.

Степень внедрения: средняя, продукт представляет из себя готовый алгоритм, но предстоит его доработка. Область применения: полученные результаты могут быть использованы сельскохозяйственными компаниями, в частности компаниями АПК сектора.

Бакалаврская работа выполнена в текстовом редакторе Microsoft Word, для расчетов использовались пакеты программ MS Excel, Python, R, Wolfram, Statistica.

Оглавление

РЕФЕРАТ.....	9
Введение.....	11
1. Обзор литературы.....	13
2. Объект и методы исследования.....	15
2.1 Случайный лес (Random forest).....	16
2.2 Линейная регрессия.....	17
2.3 Многослойный перцептрон (MLP).....	18
3. Практические расчеты и анализ результатов.....	20
3.1 Сопоставление данных.....	20
3.2 Анализ набора данных.....	24
3.3 Описание реализации.....	25
3.4 Анализ моделей.....	27
4. Результаты проведенного исследования.....	34
5. Раздел «Стартап как диплом».....	39
5.1 Описание продукта.....	39
5.2 Интеллектуальная собственность.....	39
5.3 Объем и емкость рынка. Анализ современного состояния и перспектив.....	40
5.4 Планируемая стоимость.....	46
5.5 Конкурентные преимущества.....	47
5.6 Целевые сегменты потребителей.....	48
5.7 Бизнес модель проекта.....	50
5.8 Итог.....	52
6. Раздел «Социальная ответственность».....	53
6.1 Введение.....	53
6.2 Правовые и организационные вопросы обеспечения безопасности.....	54
6.4 Экологическая безопасность.....	61
6.5 Безопасность в чрезвычайных ситуациях.....	63
6.6 Выводы по разделу.....	64
Заключение.....	66
Список литературы.....	67
Приложение А.....	70
Приложение Б.....	71

Введение

Реализация сельскохозяйственных работ подразумевает под собой не только определенные алгоритмы действий для посадок, но и так же анализ различных факторов, влияющих на рост урожая, эти параметры являются индивидуальными в зависимости от местности. Как следствие, с увеличением объемов проекта, увеличивается объем и обрабатываемой информации, с которой необходимо работать. Под влиянием человеческого фактора это может неизбежно привести к увеличению количества ошибок в оценке урожайности продукции.

Для того чтобы значительно сократить разброс оценки, человек может собрать как можно большее количество данных, влажность, химический состав почвы, уровень зрелости и т.д. После чего провести кропотливую работу над общим анализом. Но при подобной работе выделяются ряд недостатков:

- Человеческий фактор: всегда существует вероятность, что человек, который проводит анализ, может допустить ошибку или пропустить определенный набор данных
- Большое количество времени, затраченное на поиск проблем и составление оценки.

Есть альтернативный вариант выполнения поставленной задачи – доверить оценку машине. Но до относительно недавнего времени данному способу препятствовало ряд причин:

- Машина не была способна сама систематизировать данные по степени их важности, а при «объяснении» вручную человек затратил бы гораздо больше времени чем при анализе параметров и составлении выводов своими силами
- Недоступность достаточно мощных систем для «кустарной» обработки подобных алгоритмов

На данный момент мы обладаем необходимыми алгоритмами для выявления наиболее влиятельных факторов, независимо от предоставляемых данных, а также большая доступность вычислительных мощностей, позволяет вести расчеты не только на теоретическом уровне, но и дает возможность производить моделирование на основе существующих данных.

Исходя из вышесказанного ставится цель по решению проблемы оценки урожайности картофеля с использованием алгоритмов регрессионного анализа. Мы предлагаем использовать три метода: а) Random Forest (случайные деревья), б) Linear regression (линейная регрессия), в) MLP (многослойный перцептрон).

Для достижения поставленной цели выявлены следующие задачи:

- 1) Анализ датасета и его подготовка для дальнейшей работы и использования в обучении.
- 2) Анализ нескольких регрессионных алгоритмов, с целью выяснения наиболее оптимального для реализации проекта.
- 3) Обучение модели.
- 4) Написание кода, работающего с параметрами модели и со стартовым датасетом с целью формирования предсказанных данных.
- 5) Проведение сравнения предсказанных характеристик с реальными.

В данной работе предоставлен пример решаемой проблемы по оценке урожайности картофеля, на основе многомерного регрессионного анализа, опирающегося на параметры химического состава почвы, а также сорта и группы зрелости картофеля.

1. Обзор литературы

Общее управление растениеводствам. В целом, системы управления растениеводством обеспечивают интерфейс для общего слежения за сельскохозяйственными культурами, охватывают каждый аспект его ведения.

Идея использования искусственного интеллекта в управлении растениеводством была впервые предложена в 1985 году Маккинионом и Леммоном в статье [1]. Другая экспертная система защиты посевов кукурузы была предложена Буланже в его докторской диссертации [2]. В 1987 году Роуч и др. предложили экспертную систему РОММЕ для управления яблоневыми плантациями [3]. Стоун и Томан разработали экспертную систему для управления посевами хлопчатника COTFLEX [4]. Многослойная система на основе искусственной нейронной сети с прямой связью была разработана Робинсоном и Мортом для защиты цитрусовых культур от повреждения морозом на итальянском острове Сицилия [5]. Входные и выходные параметры были закодированы в двоичной форме, авторы использовали различные конфигурации входных данных, чтобы получить модель с наивысшей точностью. Лучшая найденная модель имела точность 94% с двумя выходными классами и шестью входными сигналами. Применялся искусственный интеллект, работающий с изображениями, для посева пшеницы [6], с использованием алгоритма маркировки пикселей с последующим преобразованием Лапласа для усиления выходной информации. Лучшая полученная сеть имела пять скрытых слоев, обученных до 300000 итераций, и имела точность в среднем 85,9%.

Основанная на нечеткой логике система управления урожаем сои была разработана Пракашем с коллегами. Данная система давала советы относительно выбора сельскохозяйственных культур, внесения удобрений и вопросов, связанных с вредителями [7].

Болезни сельскохозяйственных культур. Болезни сельскохозяйственных культур также являются предметом серьезной

озабоченности фермеров. Для того чтобы обнаружить больное растение и предпринять необходимые шаги для восстановления, требуются значительные знания и опыт. Компьютерные системы используются во всем мире для диагностики заболеваний и предложения мер борьбы.

На очень ранней стадии были разработаны системы [8-10]. Тильва с соавторами предложили модель, основанную на нечеткой логике, для прогнозирования заболеваний на основе продолжительности увлажнения листьев [11]. Различные модели, основанные на искусственной нейронной сети, были предназначены для борьбы с болезнями различных культур [12-18]. Также были предложены некоторые гибридные системы, модель обработки изображений в сочетании с моделью искусственной нейронной сети для классификации болезней рассады фаланопсиса [19]. Саннакки и др. применили подход нечеткой логики в сочетании с обработкой изображений для определения процента заражения в листьях [20]. Системы, использующие алгоритм сегментации k-средних, были предложены в [21, 22]. Отметим также экспертную веб-систему Dr. Wheat для диагностики заболеваний пшеницы [23].

Контроль хранения и качества сельхозпродукции. Помимо мониторинга вредителей и болезней, хранение, сушка, сортировка собранных культур также являются очень важными аспектами сельского хозяйства. В этом разделе рассматриваются различные механизмы мониторинга пищевых продуктов и контроля качества, использующие концепцию искусственного интеллекта.

Было разработано несколько систем, основанных на нечеткой логике, в работах: Мартыненко и Янг [24], Мовагарнеджад и Никзад [25], Бонецки [26].

Управление гербицидами. Применение гербицидов также оказывает прямое воздействие на здоровье человека и окружающую среду. Современные методы искусственного интеллекта применяются для минимизации применения гербицидов за счет надлежащей и точной борьбы с сорняками.

Паскуаль [27] разработал систему для выявления и устранения сорняков в таких культурах, как овес, ячмень, тритикале и пшеница. Беркс [28] использовал машинное зрение с обученной обратному распространению нейронной сетью для идентификации сорняков пяти различных видов. Беркс с коллегами [29] сравнили три различные модели нейронных сетей, главным образом обратное распространение, встречное распространение и модель на основе радиальной базисной функции с тем же набором входных данных, что и в предыдущей статье, и обнаружили, что сеть обратного распространения лучше всего работает с точностью 97%.

Прогнозирование урожайности. Прогнозирование урожайности сельскохозяйственных культур очень полезно для маркетинговых стратегий и оценки стоимости урожая. Более того, в эпоху точного земледелия анализ соответствующих факторов, непосредственно влияющих на урожайность, также может быть проведен с помощью моделей прогнозирования.

Лю с коллегами [30] использовали модель искусственной нейронной сети, использующую алгоритм обучения обратному распространению, для прогнозирования урожайности по параметрам почвы. Эрет с авторами построил нейронную модель для прогнозирования урожайности томатов, их роста и водопотребления в тепличных условиях [31]. Тонгбуннак и Сарапироме экспериментировали с урожайностью пшеницы в разных районах Таиланда, используя нейронные сети [32]. В другом подходе Пехлеван и коллеги использовали выработку энергии в качестве показателя урожайности растений базилика в теплице [33]. Другие важные результаты прогнозирования урожайности представлены в [34-36].

В 2014 году Дахикар и Роде предложили нейронную модель для прогнозирования урожайности 7 различных культур с использованием атмосферных воздействий и потребления удобрений [37].

2. Объект и методы исследования

Объектом исследования является набор данных института Квебека, Канады «Department of Soils and Agrifood Engineering, Université Laval, Québec,

QC, Canada, 2 Department of Civil and Water Engineering, Université Laval, Québec, QC, Canada» (Приложение А).

Перед выполнением поставленной задачи, необходимо провести теоретический анализ предметной области с целью пояснения рабочих моментов. Начать следует с пояснений к работе обучающих алгоритмов

2.1 Случайный лес (Random forest)

Это ансамблевый алгоритм машинного обучения, который используется для задач классификации и регрессии. Он строит несколько деревьев решений и объединяет их для получения более точного и стабильного прогноза. Random Forest менее склонен к переобучению и может обрабатывать большие наборы данных с высокой размерностью.

Рассмотрим обучающий набор данных с выборками и признаками, представленный матрицей X , и целевую переменную y , представленную вектором y .

1. Создать B деревьев решений путём бутстраппинга обучающих данных (выборка с заменой)
2. Для каждого дерева решений b :
 - a. Выбрать случайным образом k признаков из m признаков.
 - b. Найти наилучшее разбиение для каждого выбранного признака и выбрать разбиение, которое минимизирует *остаточную сумму квадратов* (RSS) для целевой переменной y .
 - c. Выращивание дерева до максимального размера без обрезки
3. Сделать прогноз для нового входного x :
 - a. Пропустить входные данные через все B деревьев решений и получить B предсказаний для y .
 - b. Вычислить среднее значение прогнозов B в качестве окончательного прогноза для y .

В математической нотации прогноз для нового входного x с помощью регрессии случайного леса может быть выражен так:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B f_b(x),$$

где \hat{y} – предсказанное значение целевой переменной, $f_b(x)$ – предсказание b -го дерева решений для входа x , B – количество деревьев решений в лесу.

Каждое дерево решений b обучается на бутстрапном наборе данных X_b и соответствующем целевом векторе y_b . Предсказание b -го дерева решений может быть выражено как:

$$f_b(x) = \sum_{j=1}^{J_b} c_{bj} I(x \in R_{bj}),$$

где J_b – количество терминальных узлов в b -ом дереве решений, c_{bj} – значение предсказания в узле j -го дерева решений, а R_{bj} – область пространства признаков, определяемая j -ым терминальным узлом.

Критерием разбиения, используемым в каждом дереве решений, обычно является *средняя квадратичная ошибка* (MSE), задаваемая следующим образом:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2,$$

где N – количество образцов в обучающих данных, y_i – целевая переменная для образца i , а \hat{y} – предсказанное значение для y_i . Для каждого выбранного признака выбирается разбиение, минимизирующее MSE.

2.2 Линейная регрессия

Статистический метод, используемый для анализа взаимосвязи между двумя переменными, а именно зависимой переменной и одной или несколькими независимыми переменными. Он предполагает, что между переменными существует линейная связь и пытается найти наилучшую линию, которая может предсказать значение зависимой переменной на основе независимых переменных.

Учитывая набор данных с n наблюдениями зависимой переменной y и независимой переменной x , модель линейной регрессии стремится оценить связь между y и x путем обучения линии вида:

$$y = \beta_0 + \beta_1 x + \epsilon,$$

где β_0 – смещение (значение y при $x=0$), β_1 – наклон (изменение y на единицу изменения x), и ϵ – член ошибки (представляющий случайную изменчивость y , не объясняемую x).

Параметры β_0 и β_1 могут быть оценены с помощью метода наименьших квадратов, который стремится минимизировать сумму квадратов между наблюдаемыми значениями y и предсказанными значениями из модели:

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 \rightarrow \min,$$

где \hat{y}_i – предсказанное значение y для i -го наблюдения, определяемое:

$$\hat{y}_i = \beta_0 + \beta_1 x.$$

Оценки β_0 и β_1 , которые минимизируют эту сумму, даются следующим образом:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x},$$

где \bar{x} и \bar{y} – выборочные средние x и y , соответственно.

Таким образом, окончательная модель простой линейной регрессии может быть записана как:

$$y = \hat{\beta}_0 + \hat{\beta}_1 x + \epsilon.$$

2.3 Многослойный перцептрон (MLP)

Это тип нейронной сети, которая используется для задач контролируемого обучения. Она состоит из входного слоя, одного или нескольких скрытых слоев и выходного слоя. Каждый нейрон в сети связан с нейронами в соседнем слое, и каждое соединение имеет вес, связанный с ним. Сеть обучается, регулируя эти веса посредством обратного распространения.

Учитывая набор данных с n выборками и m входными признаками, целью MLP является обучение функции $f(x)$, которая отображает вход x в R^m

на выход y в R^k , где k – количество выходных классов или целей регрессии. MLP состоит из одного или нескольких скрытых слоев нейронов, каждый нейрон берет взвешенную сумму своих входов и пропускает ее через функцию активации

$$z_i^{(l)} = \sum_{j=1}^{m^{(l-1)}} \omega_{ij}^{(l)} x_j^{(l-1)} + b_i^{(l)},$$

$$a_i^{(l)} = \sigma(z_i^{(l)}),$$

где i – индекс нейрона в слое l , j – индекс входного признака, $\omega_{ij}^{(l)}$ – вес связи между входом j и нейроном i в слое l , $b_i^{(l)}$ – член смещения для нейрона i в слое l , $m^{(l-1)}$ – количество нейронов в предыдущем слое, $\sigma(\cdot)$ – функция активации, а $z_i^{(l)}$ и $a_i^{(l)}$ – взвешенная сумма и активация нейрона i в слое l , соответственно.

Выходной сигнал MLP задается:

$$y = f(x) = g\left(\sum_{j=1}^{m^{(L-1)}} \omega_{ij}^{(L)} a_j^{(L-1)} + b_i^{(L)}\right)$$

где L – индекс выходного слоя, $\omega_{ij}^{(L)}$ – веса, соединяющие последний скрытый слой с выходным слоем, $b_i^{(L)}$ – член смещения для j -го выходного узла, $g(\cdot)$ – функция активации для выходного слоя, которая может отличаться от функций активации, используемых в скрытых слоях, в зависимости от задачи.

Параметры MLP, включая веса и смещения, изучаются путем минимизации функции потерь $L(y, \hat{y})$ между прогнозируемым выходом \hat{y} и истинным выходом y с помощью обратного распространения и градиентного спуска.

Для оптимизации производительности перечисленных выше моделей использовался GridSearchCV. GridSearchCV – метод настройки гиперпараметров, который исчерпывающим образом перебирает заранее определенный набор гиперпараметров, чтобы найти наилучшую комбинацию

гиперпараметров, оптимизирующую производительность модели. В приведенном коде GridSearchCV использовался для поиска наилучшего набора гиперпараметров для каждой модели.

Далее приступаем к практической реализации алгоритма. Начать следует с анализа набора данных и его климатического соответствия регионам России.

3. Практические расчеты и анализ результатов

3.1 Сопоставление данных.

Кратко опишем регион Квебек с целью проведения аналогий с регионами России. Квебек имеет три основных климатических региона.

1) На юге и западе региона Квебек преобладает влажный континентальный климат. Данной области присущи теплое лето и холодная зима. Осадки в данном регионе достигают более 1000 мм, из-за того, что регион находится под влиянием ветров Северной Америки и Атлантического океана. Слой снега зимой может достигать до 300 см в высоту. Подобный континентальный климат характерен для Западной Сибири и юго-востока Европейской равнины. Таким образом к данной области можно отнести такие города как Томск, Новосибирск, Абакан, Москва и т.д.

2) Центральные районы Квебека относятся к зоне субарктического климата. Характерны холодные и длинные зимы и теплое, но короткое лето, по причине сильного влияния арктических воздушных масс. Также стоит отметить, что в данном регионе число выпадения осадков составляют значительно меньшие объемы по сравнению с южным районом. В России субарктический тип климата имеют территории, находящиеся за полярным кругом в области Восточно-Европейской равнины, а также Западной Сибири. Так же данный тип климата преобладает на всех территориях северо-востока РФ к северу от 60° с.ш. К данной области можно отнести такие города как Якутск, Иркутск, Магадан.

3) Север области Квебек составляет зона арктического климата, зимы холодные, а лето самое короткое в регионе. Наиболее важное влияние ветрами на регион оказывают течения Северного Ледовитого океана и континентальные воздушные массы Арктики. Здесь можно отметить аналогии с сибирским побережьем Северного Ледовитого океана и прилегающие острова, за исключением южного острова Новой Земли, островов Вайгача, Колгуева и других в южной части Баренцева моря. К данной области можно отнести такие города как Нарьян-Мар, Норильск, Магадан.

Как было сказано выше исследования были проведены для интервала широт [45, 49] с акцентом на правую границу, где располагается исследовательский центр – Лаваль в составе мегаполиса Большой Монреаль в провинции Квебек, Канада.

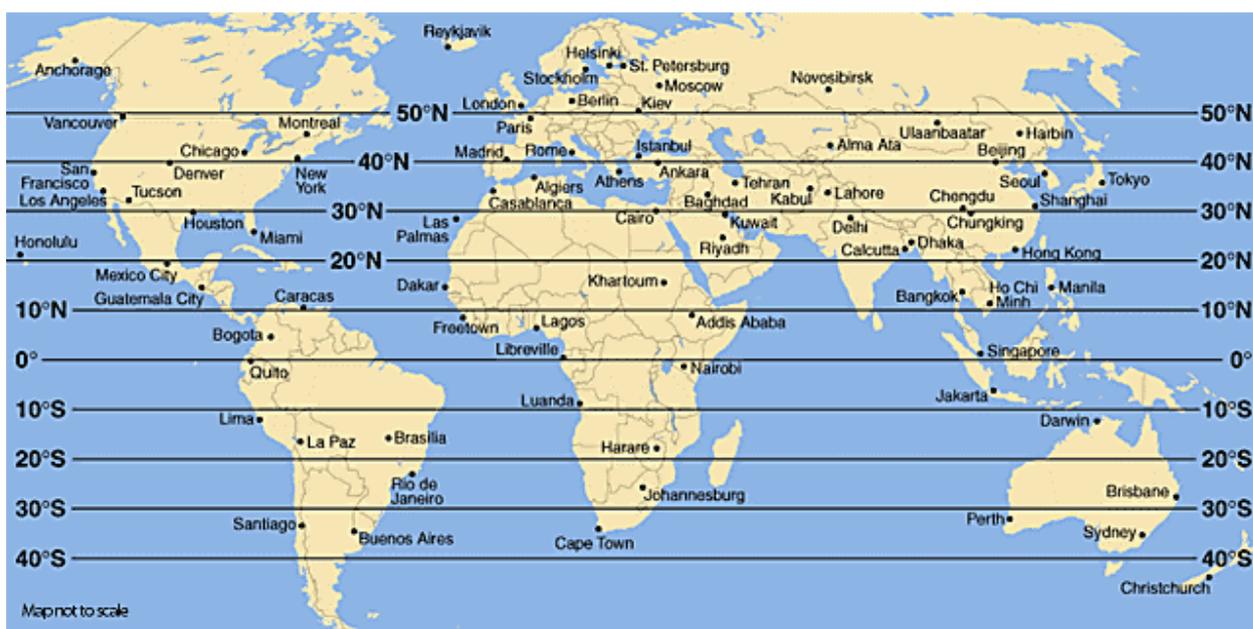


Рис. 1. Сопоставление широт Канады и России

Климат Монреалья ($45^{\circ}30'32''$ с. ш. $73^{\circ}33'15''$ з. д.) соответствует влажному континентальному, данный тип климата классифицируется буквой «D». Так как средняя высокая температура в городе летом превышает 22°C , что соответствует жаркому лету, но при этом среднее минимальное число осадков и зимой, и летом более 30 мм, можно сделать вывод о трудности

чёткой классификации Монреаля, данный тип климата помимо буквы «D», добавляет приписку «Dfb»

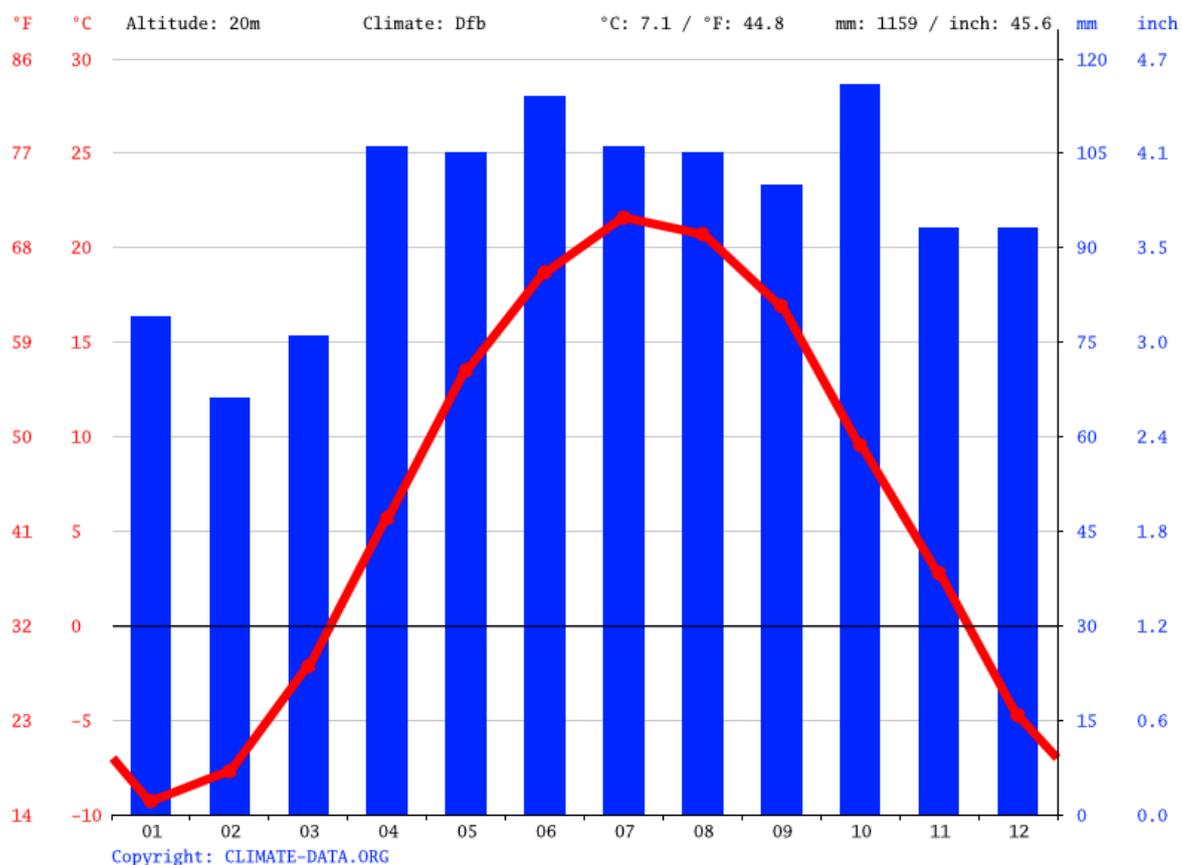


Рис. 2. Климатический портрет Монреаля

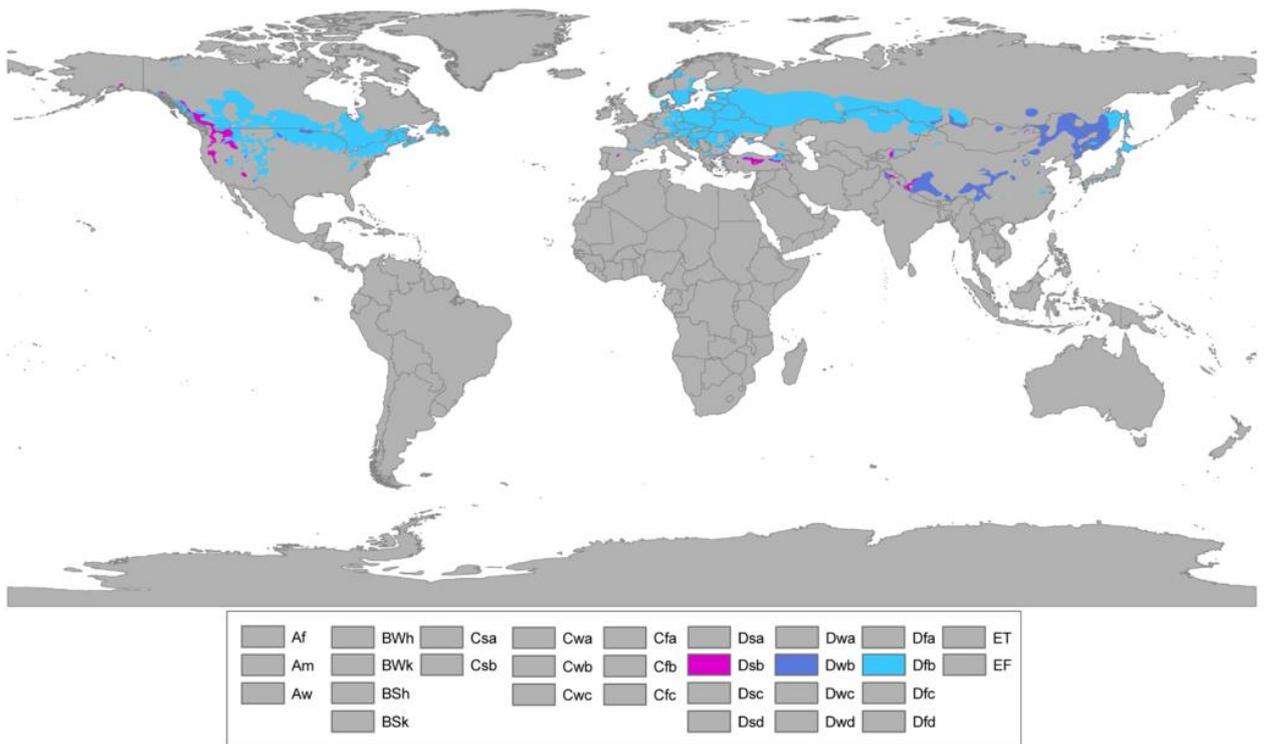


Рис. 3 Климатические зоны по Кёппену

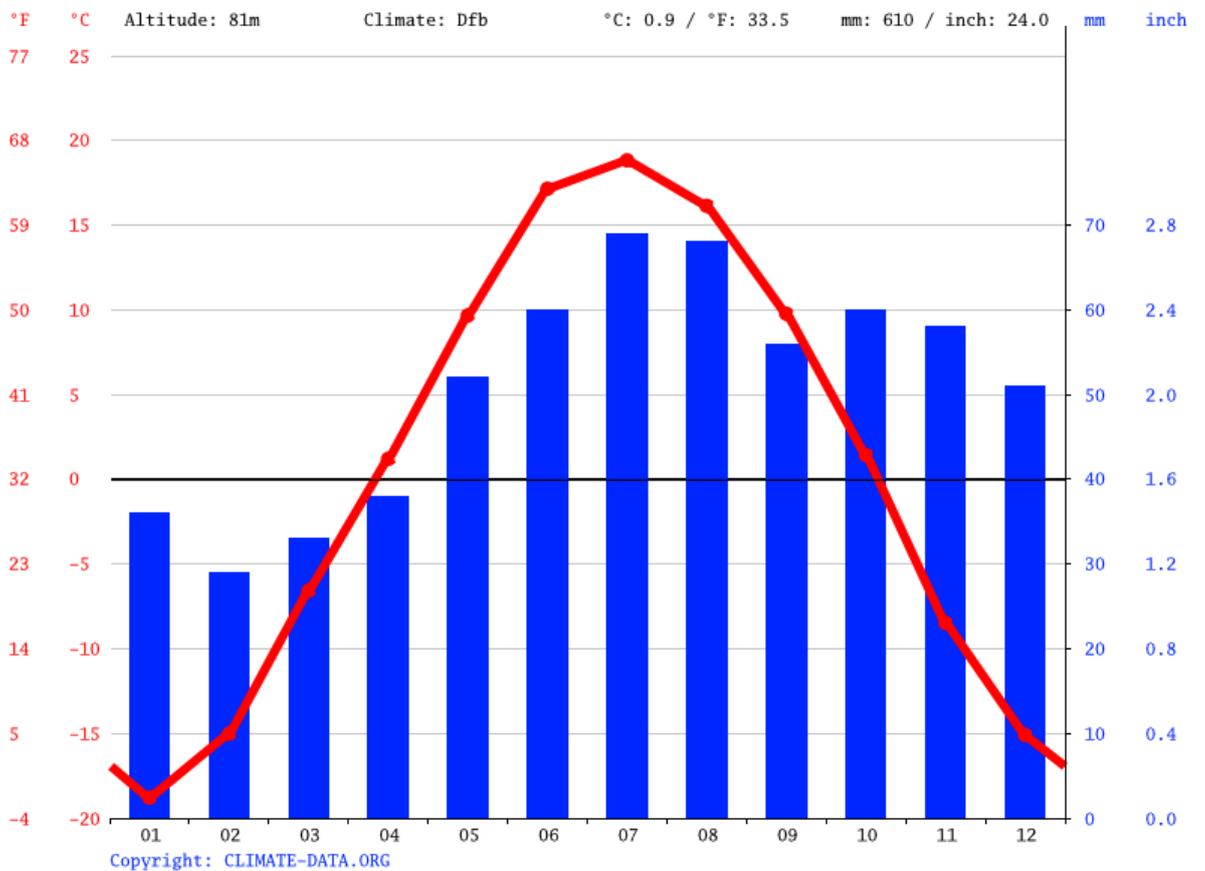


Рис. 4. Климатический портрет города Томск

Как мы видим регионы с похожим климатом, именуемым как «влажный континентальный климат с тёплым летом», находятся большей части западной России, а также в южной части Сибири. Это такие города как Москва, Санкт-Петербург, Казань, Красноярск, Новосибирск, Томск

После проведения климатического анализа можно приступать к анализу исходного набора данных.

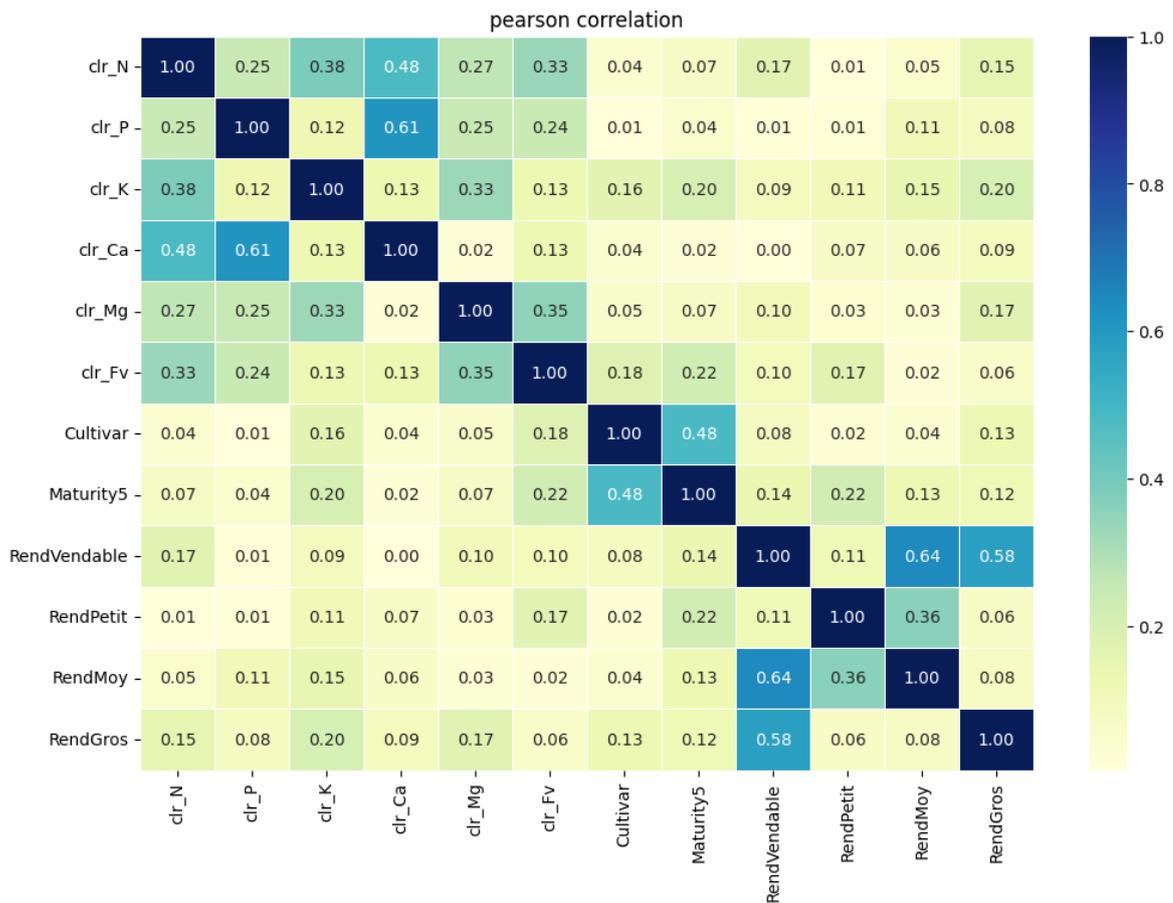
3.2 Анализ набора данных.

Датасет «leaf_clust» (Приложение 1) представляет из себя набор данных, содержащих в себе такие параметры как:

- N, P, K, Ca и Mg compositions = макроэлементы (выражены в процентах)
- Fv = filling value = $100 - \text{sum}(N, P, K, Ca, Mg)$
- Maturity5 = группы зрелости картофеля
- RendVendable = товарная доходность
- RendPetit = товарная доходность картофеля большого размера
- RendMoy = товарная доходность картофеля среднего размера
- RendGross = товарная доходность картофеля малого размера
- Cultivar = сорт
- NoEssai = номер испытания
- NoBloc = номер объединения
- NoTraitment = номер процедуры

Стоит отметить, что характеристики Petit, Moy и Gros являются характеристиками размера картофеля (большой средний и малый), и если их просуммировать, то итог дает характеристику Vendable. Данный набор из четырёх переменных является результирующим и именно данные параметры необходимо использовать для оценки предсказанной урожайности

В качестве предикторов выбираем параметры химического состава почвы, а та же ещё две характеристики (группа зрелости и сорт). Остальные характеристики представляют собой или описательные признаки или выводятся из признаков, выбранных выше.



Корреляции (Лист1 в Analys) Отмеченные корреляции значимы на уровне $p < 0,05000$ N=3111 (Построчное удаление ПД)												
Переменная	Средние	Ст.откл.	clr_N	clr_P	clr_K	clr_Ca	clr_Mg	clr_Fv	RendPetit	RendMoy	RendGros	RendVendable
clr_N	0,74126	0,17878	1,000000	0,271456	-0,303964	-0,421511	-0,243286	0,069630	0,006606	0,049735	0,155449	0,150810
clr_P	-2,00832	0,24273	0,271456	1,000000	-0,359170	-0,590337	-0,151504	-0,071556	-0,011675	0,109271	-0,082759	0,033029
clr_K	0,44441	0,27470	-0,303964	-0,359170	1,000000	0,008261	-0,567493	0,220631	0,111107	-0,145432	0,205548	0,064126
clr_Ca	-1,00896	0,23751	-0,421511	-0,590337	0,008261	1,000000	0,022542	-0,220694	-0,065536	0,062086	-0,092818	-0,039857
clr_Mg	-1,72138	0,28674	-0,243286	-0,151504	-0,567493	0,022542	1,000000	-0,447409	0,028432	-0,031242	-0,173075	-0,128954
clr_Fv	3,55300	0,12502	0,069630	-0,071556	0,220631	-0,220694	-0,447409	1,000000	-0,171615	-0,010017	0,060041	-0,049202
RendPetit	6,77295	5,51954	0,006606	-0,011675	0,111107	-0,065536	0,028432	-0,171615	1,000000	-0,356008	-0,063675	0,115457
RendMoy	20,93108	10,22477	0,049735	0,109271	-0,145432	0,062086	-0,031242	-0,010017	-0,356008	1,000000	-0,080636	0,642198
RendGros	6,00426	8,01484	0,155449	-0,082759	0,205548	-0,092818	-0,173075	0,060041	-0,063675	-0,080636	1,000000	0,577581
RendVendable	33,69313	11,86036	0,150810	0,033029	0,064126	-0,039857	-0,128954	-0,049202	0,115457	0,642198	0,577581	1,000000

Рис. 5.6. Матрицы корреляций основных характеристик

После анализа можно приступать к формированию алгоритма реализации.

3.3 Описание реализации.

Приведем пошаговое описание реализации алгоритма:

- Считывание CSV файла leaf_clust_df.csv и выбор числовых признаков, начинающихся с "clr", добавление столбцов "Cultivar" и "Maturity5" вместе с возможными целевыми переменными ("RendVendable", "RendPetit", "RendMoy", "RendGros") для создания нового DataFrame.

Создание тепловой карты корреляций для вновь созданного DataFrame с помощью функции `plot_corr`.

- Удаление из DataFrame всех строк с нулевыми значениями.
- Создание гистограммы для целевых переменных и некоторых других характеристик DataFrame.
 - Добавление в DataFrame one-hot-encoded переменных для столбцов "Maturity5" и "Cultivar".
 - *for {x} in {"RendVendable", "RendPetit", "RendMoy", "RendGros"} select x as Target.*
 - Разделение DataFrame на обучающий и тестовый наборы с помощью функции `train_test_split` из `sklearn.model_selection`.
 - Масштабирование числовых характеристик обучающего и тестового наборов с помощью `StandardScaler` из `sklearn.preprocessing`.
 - Определение трех моделей регрессии - Random Forest Regressor, Linear Regression и Multi-Layer Perceptron, используя `RandomForestRegressor`, `LinearRegression` и `MLPRegressor` из `sklearn.ensemble`, `sklearn.linear_model` и `sklearn.neural_network` соответственно.
 - Определение сетки гиперпараметров для поиска по каждой модели.
 - Создание объектов `GridSearchCV` для каждой модели с использованием заданных сеток гиперпараметров.
 - Обучение `GridSearchCV` на обучающих данных.
 - Вывод лучших гиперпараметров и лучших оценок для каждой модели.
 - Оценка лучшей модели на тестовых данных и вычисление средней квадратичной ошибки для каждой модели.
 - Построение гистограммы важности признаков для модели, показавшей наилучшие результаты
 - Построение графика фактических и прогнозируемых значений для лучшей модели на тестовых данных.

На данном этапе набор данных уже подготовлен для обучения. Таким образом, в работе открывается возможность обучения моделей для последующего анализа их эффективности.

3.4 Анализ моделей

Сравнение регрессионных моделей для прогнозирования RendVendable.

Регрессионная модель Random Forest превзошла две другие модели по всем показателям оценки. Она достигла самого высокого показателя R^2 , равного 0.6, и самых низких значений MSE и RMSE - 0.18 и 8, соответственно. Она также имела самую низкую максимальную ошибку 36, что указывает на ее лучшую способность предсказывать экстремальные значения.

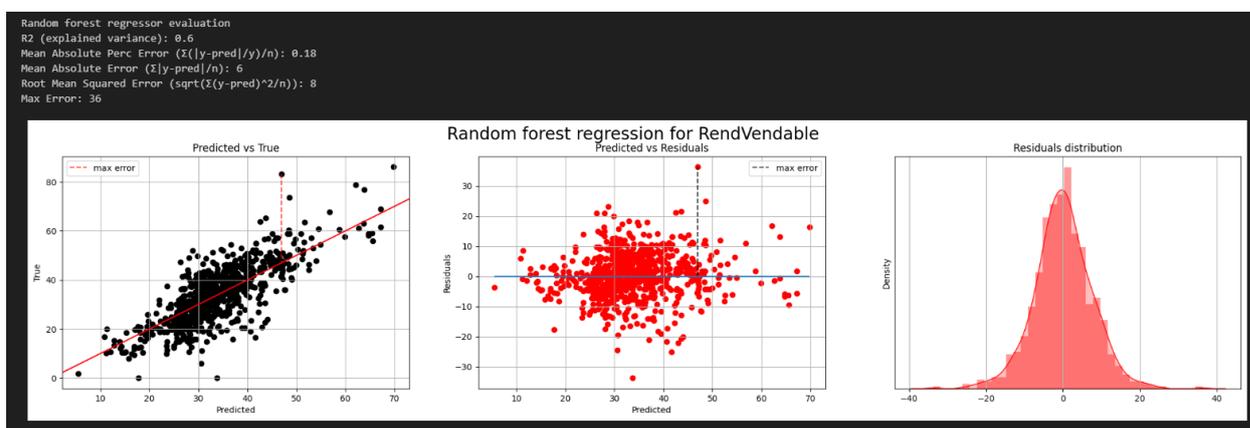


Рис. 7 «Характеристики модели RandomForest»

Модель линейной регрессии имела самый низкий показатель R^2 - 0.23 и самые высокие значения MSE и RMSE - 0.23 и 11 соответственно. Так же модель имеет одну из самых высоких минимальных ошибок - 50, а значит самое худшее предсказание экстремальных значений

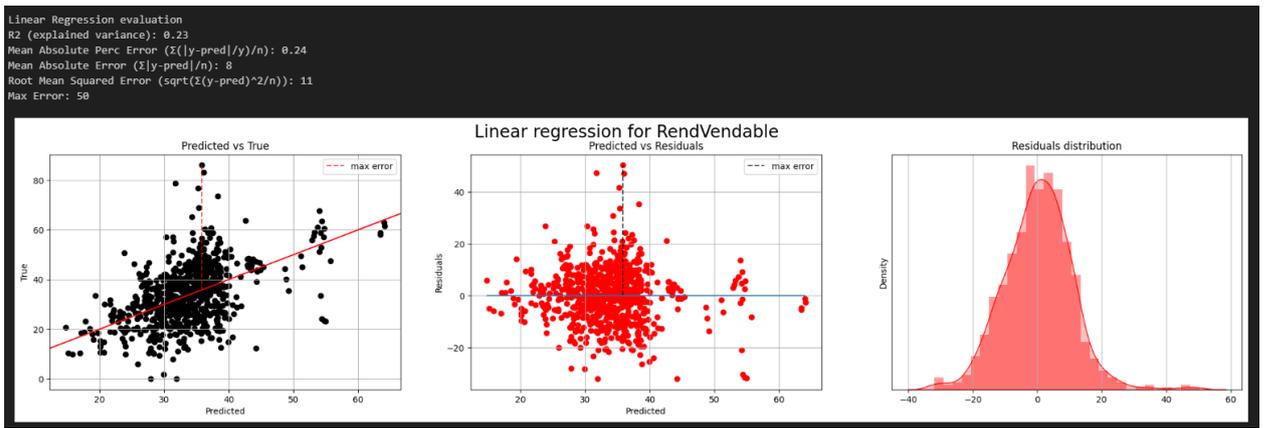


Рис. 8 «Характеристики модели LinearRegression»

Модель многослойного перцептрона имеет показатель $R^2=0.46$, что находится между показателями моделей регрессии случайного леса и линейной регрессии. Ее значения MSE и RMSE также находились между значениями двух других моделей. То же самое можно сказать про меньшую максимальную ошибку - 49.

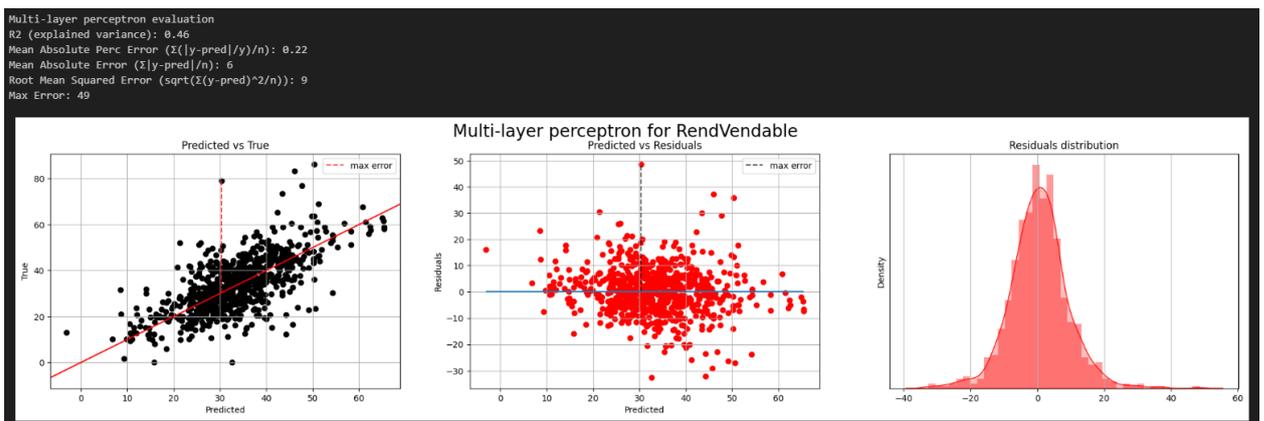


Рис 9. «Характеристики модели Multi-layer perceptron»

В целом, модель регрессии случайного леса кажется лучшей среди трех моделей по данным показателям оценки

Сравнение регрессионных моделей для прогнозирования RendPetit

Модель случайного леса демонстрирует хорошие результаты с показателем $R^2=0.74$ и относительно низкими значениями MAE и RMSE. Так же низким является и показатель MAPE, что говорит о том, что она может быть достаточно точной в прогнозировании значений, это означает, что у него самое

низкое среднее отклонение между предсказанными и фактическими значениями

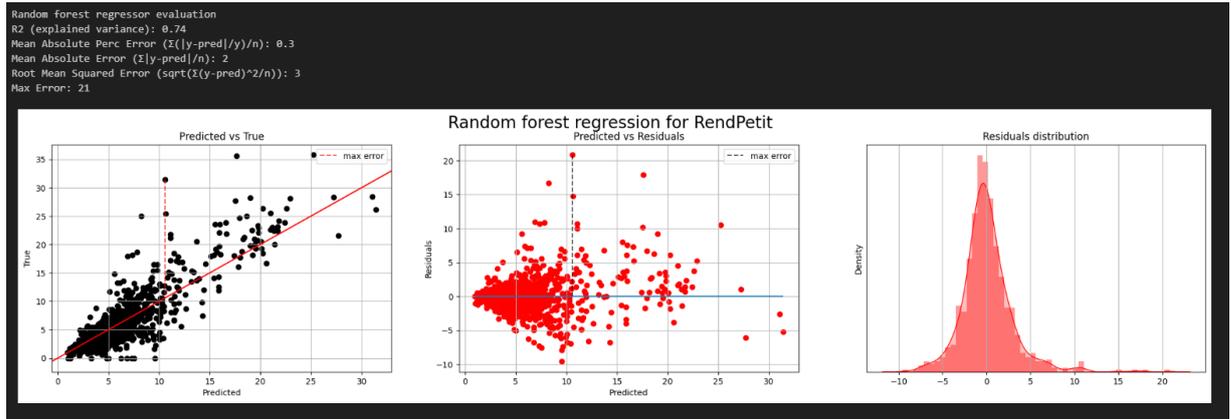


Рис. 10 «Характеристики модели RandomForest»

Модель многослойного перцептрона, похоже, показывает наилучшие результаты с точки зрения метрик оценки. У нее самый высокий показатель $R^2=0.77$, это указывает на то, что она лучше других моделей объясняет дисперсию в данных.

Средняя абсолютная ошибка в процентах (MAPE) также достаточно низкая для многослойного перцептрона (0.39). Кроме того, средняя абсолютная ошибка (MAE) и средняя квадратичная ошибка (RMSE) равны параметрам RandomForest, что говорит о наименьшей средней разнице между прогнозируемыми и фактическими значениями.

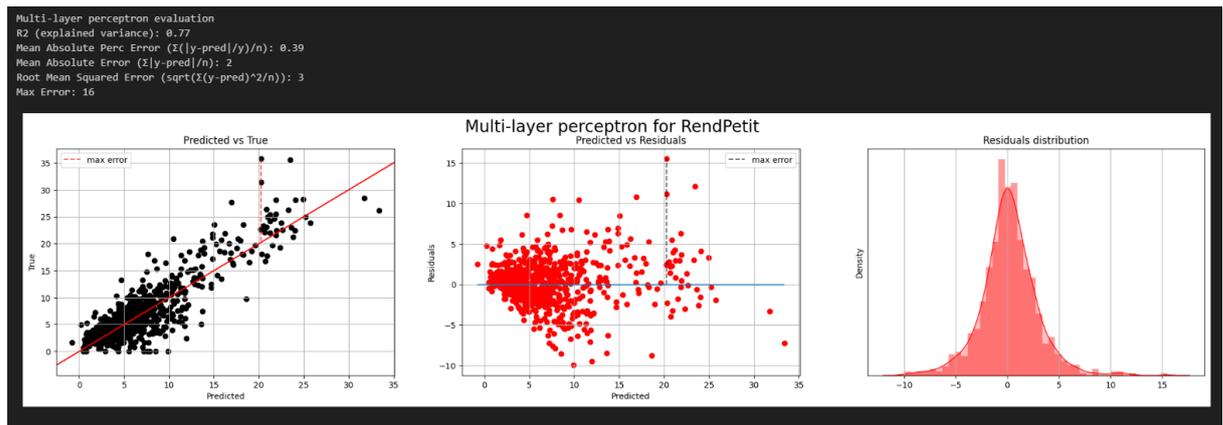


Рис. 11 «Характеристики модели Multi-layer perceptron»

Модель линейной регрессии имеет самый низкий показатель R^2 из трех моделей, что указывает на то, что она объясняет меньшую часть дисперсии в данных. Она также имеет самые высокие показатели MAPE, MAE и RMSE, что

указывает на большее среднее отклонение между прогнозируемыми и фактическими значениями.

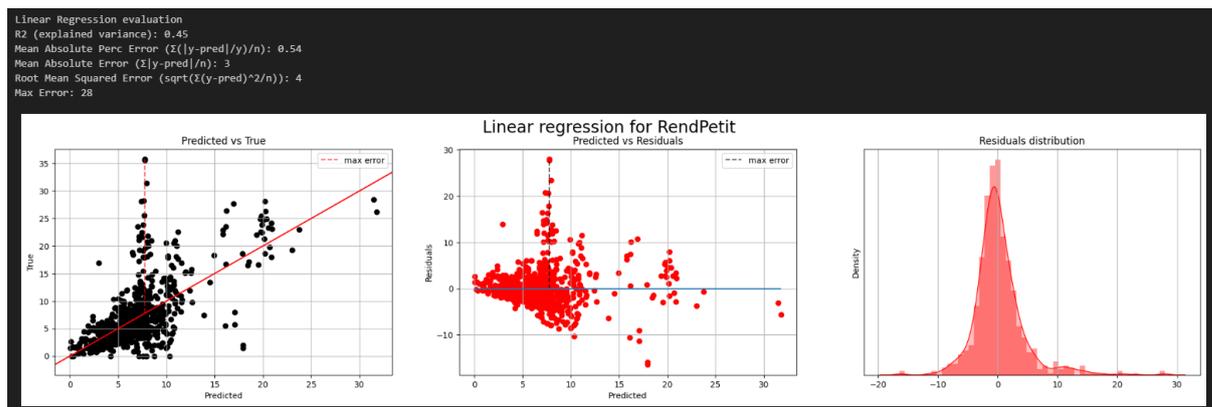


Рис. 12 «Характеристики модели LinearRegression»

В целом, модель многослойного перцептрона и модель случайного дерева подходят одинаково для реализации предсказания именно этих данных.

Сравнение регрессионных моделей для прогнозирования RendMoy

Регрессор случайного леса имеет самый высокий показатель $R^2=0.65$, что означает, что он объясняет 65% дисперсии целевой переменной. Он также имеет самую низкую среднюю абсолютную ошибку 5, что означает, что средняя разница между прогнозируемыми и фактическими значениями составляет 5 единиц. Кроме того, максимальная ошибка относительно низкая – 26, это говорит о том, что модель не делает больших ошибок в конкретной точке данных.

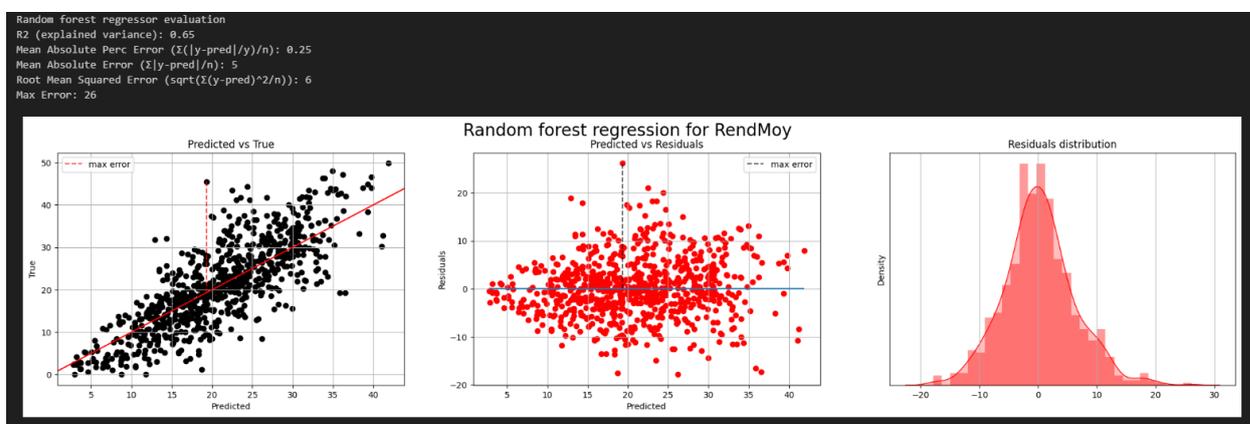


Рис. 13 «Характеристики модели RandomForest»

Многослойный перцептрон имеет более низкий показатель R^2 - 0.63, что указывает на то, что он объясняет 63% дисперсии целевой переменной. Средняя абсолютная ошибка составляет 5, что соответствует наименьшему значению у случайного леса, а максимальная ошибка самая низкая - 24. Однако средняя абсолютная процентная ошибка довольно высока - 0.29, что указывает на то, что прогнозы модели не соответствуют истинному значению в среднем на 29%.

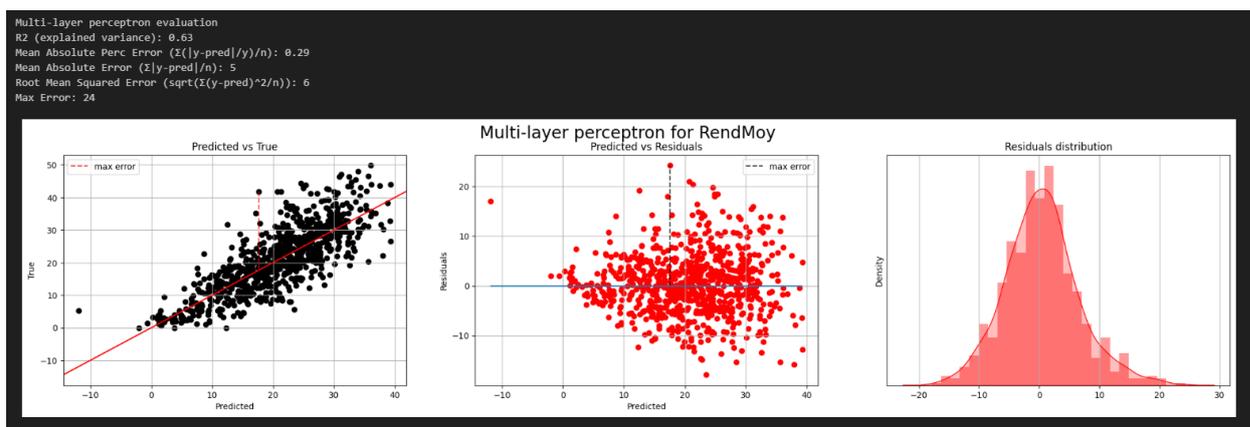


Рис. 14 «Характеристики модели Multi-layer perceptron»

Модель линейной регрессии имеет самый низкий показатель R^2 - 0.42, что означает, что она объясняет только 42% дисперсии целевой переменной. Она имеет среднюю абсолютную ошибку 6, что выше, чем у случайного леса и многослойного перцептрона, и максимальную ошибку 26, что равно случайному лесу, но больше, чем у многослойного перцептрона. Средняя абсолютная процентная ошибка также довольно высока - 0.33.

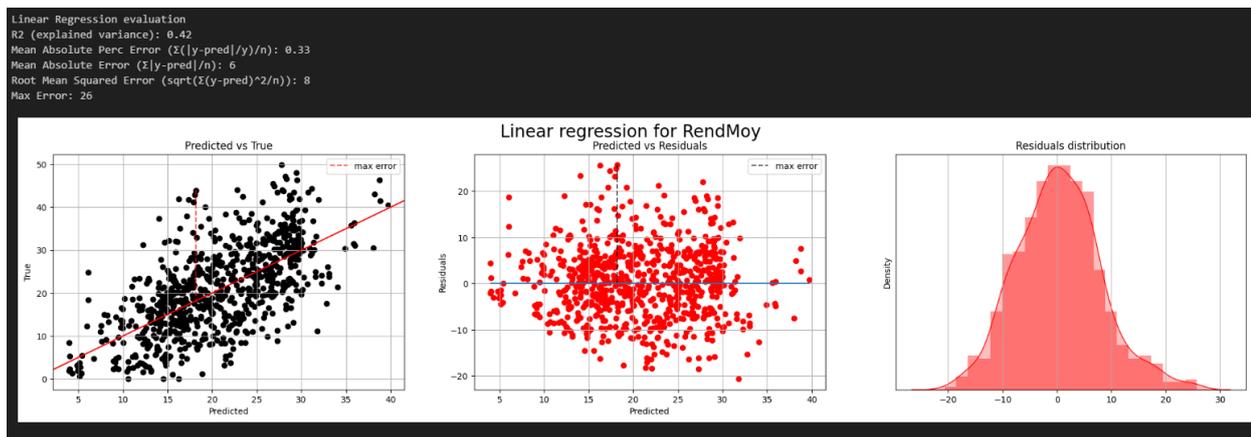


Рис. 15 «Характеристики модели LinearRegression»

В целом, и регрессор случайный лес, и многослойный перцептрон хорошо справляются с этой задачей по сравнению с линейной регрессией, при этом случайный лес имеет небольшое преимущество по средней абсолютной ошибке и R^2 , а многослойный перцептрон имеет более низкий показатель максимальной ошибки, но более высокую среднюю абсолютную процентную ошибку. Стоит отметить, что выбор лучшей модели в конечном итоге зависит от конкретных потребностей и целей решаемой задачи.

Сравнение регрессионных моделей для прогнозирования RendGros

Модель регрессора случайного леса имеет показатель R^2 , равный 0.68, что указывает на то, что она объясняет 68% дисперсии целевой переменной. Средняя абсолютная ошибка составляет 3, что соответствует показателю многослойного перцептрона и выше, чем у предыдущего результата случайного леса. Максимальная ошибка отрицательна -21, что указывает на то, что модель немного завышает предсказание на высоком уровне.

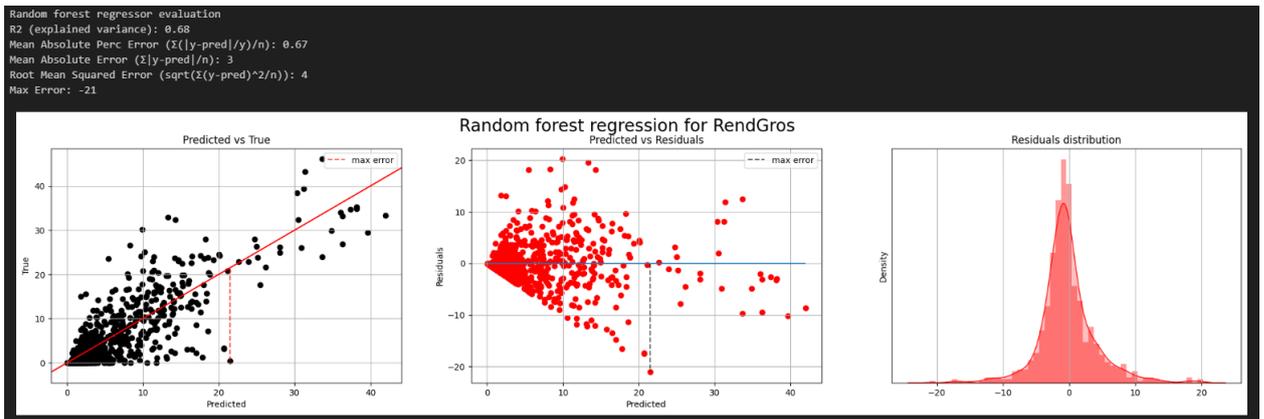


Рис. 16 «Характеристики модели RandomForest»

Модель многослойного перцептрона имеет показатель $R^2=0.6$, что ниже, чем у случайного леса. Средняя абсолютная ошибка составляет 3, что соответствует модели случайного леса, но выше, чем у предыдущей модели многослойного перцептрона. Максимальная ошибка отрицательна и составляет -32, что указывает на проблему чрезмерного предсказания, схожую с проблемой случайного леса, но в больших масштабах.

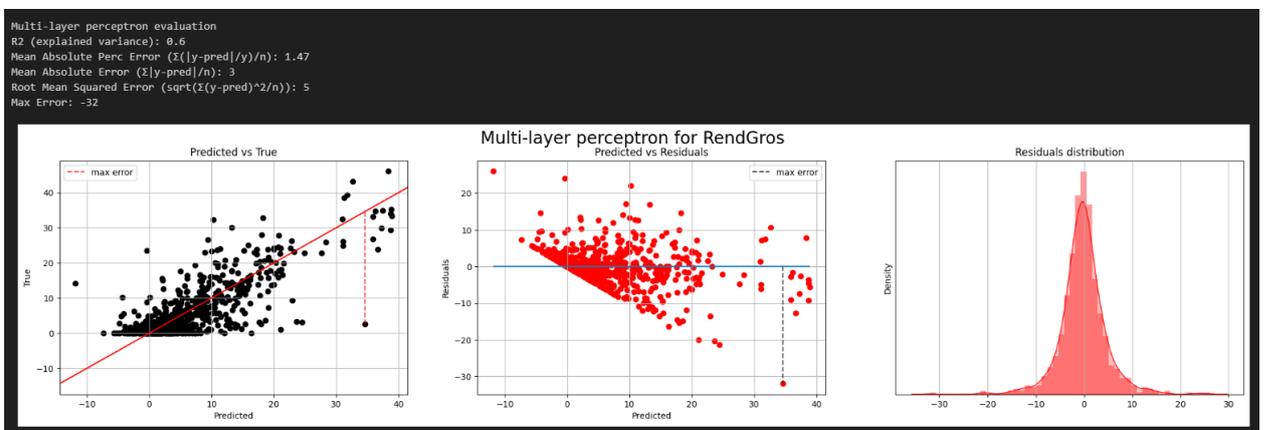


Рис. 17 «Характеристики модели Multi-layer perceptron»

Напротив, модель линейной регрессии имеет самый низкий показатель $R^2=0.33$, что указывает на то, что она объясняет только 33% дисперсии целевой переменной. У нее самая высокая средняя абсолютная ошибка - 4, а максимальная ошибка - 29, это указывает на то, что прогнозы модели менее точны, чем у модели случайного леса, но точнее многослойного перцептрона, на более высоких значениях.

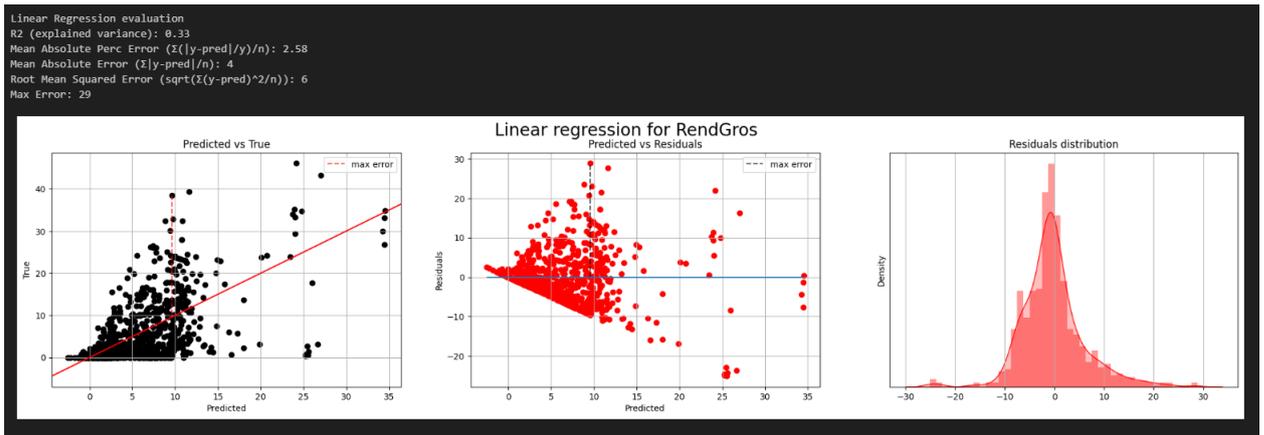


Рис. 18 «Характеристики модели LinearRegression»

В целом, модели случайного леса и многослойного перцептрона по-прежнему превосходят линейную регрессию по точности и объяснению дисперсии целевой переменной, но в данном случае предпочтение советуется отдать модели случайных деревьев

4. Результаты проведенного исследования

По окончании обучения моделей, необходимо провести сохранение всех признаков. Структурно было сохранено 12 дампов памяти, по 3 модели на каждый вид зависимой переменной.

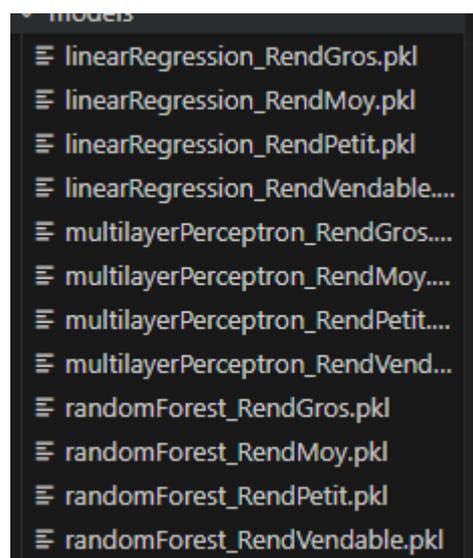


Рис. 19 «Итоговые дампы с параметрами»

Исходя из исследования эффективности моделей для каждого типа выбираем необходимый набор характеристик, предоставляемый конкретной моделью. В случае данной работы вся информация может быть предсказана исключительно с помощью результата алгоритма RandomForest.

После проведения подготовки, можно выгружать файл в котором должны содержаться такие параметры как:

- Химический состав почвы (clr параметры)
- Классификация сорта в числовом формате
- Группа зрелости так же в числовом формате
- При желании можно загрузить столбец номера измерения

Данные для загрузки перед работой стоит центрировать (Приложение Б).

Как итог на выходе программы получается таблица с предсказанными данными

X	Predict				Sum	Delta
	RendVendable	RendPetit	RendMoy	RendGros		
7947	52,80680741	4,597945813	16,81543666	34,9835414	56,39692387	3,590116457
2523	44,47558565	7,002925302	32,03039096	3,056526413	42,08984267	-2,385742977
1636	31,29673219	3,946697281	27,24508561	7,080245737	38,27202863	6,975296436
8037	64,44363571	7,387228713	19,75397432	35,75056339	62,89176642	-1,551869294
1273	30,81114986	3,217669439	27,96284572	1,842464428	33,02297959	2,211829722
266	28,07510866	6,927607681	15,43322985	5,596609115	27,95744665	-0,117662015
7909	37,38134492	8,23296479	13,87110532	14,84493994	36,94901004	-0,43233488
1882	42,08387032	11,47424388	22,23192051	7,562512461	41,26867685	-0,81519347
2542	29,65073683	8,37843523	16,5323875	6,230381609	31,14120434	1,490467504
1838	48,59612727	7,418194974	23,51884809	17,17799614	48,11503921	-0,481088059
8089	43,81358505	5,860726122	16,7553652	24,80538088	47,42147221	3,60788716
871	24,0205239	5,537474497	14,86086579	7,243120586	27,64146087	3,620936968

Рис. 20 «Оценка предсказанных данных»

В оценочной таблице можем наблюдать ряд параметров:

- X – номер конкретного измерения, отсылающий к измерению из таблиц *X_test* и *leaf_clust_df*
- Rend – характеристики – предсказанные параметры урожайности картофеля в кг
- Sum – суммарное значение характеристик Petit, Moy, Gros. Данная характеристика вызвана с целью сравнения характеристики с Vendable, ведь

исходя из изначальных данных характеристика Vendable, есть составляющая из сумм Petit, Moy, Gros.

- Delta – Разница между характеристиками Vendable и Sum

Далее сравним полученные значения с аналогичными изначальными из *leaf_clust_df*, по соответствию с переменной X

X	Real			
	RendVendable	RendPetit	RendMoy	RendGros
7947	67,59	5,06	16,43	46,1
2523	52,12527964	5,108277405	45,66174497	1,355257271
1636	35,35519126	2,131147541	33,22404372	0
8037	67,66	6,95	20,54	40,17
1273	13,73114754	2,926775956	10,80437158	0
266	45,30684804	4,010300564	36,57012181	4,726425664
7909	37,49	8,96	14,84	13,69
1882	47,4	8,2	24,3	14,9
2542	31,6994036	5,654383538	17,36980149	8,675218579
1838	38	6,5	19,5	12
8089	57,69	4,49	19,51	33,7
871	46,0134675	28,8513225	13,558545	3,6036

Рис. 21 «Реальные значения характеристик»

Проведем сравнение с таблицей выше, посчитав разницы соответствующих элементов, а также выделив максимальное и минимальное отклонение всей таблицы, посчитаем среднее отклонение и медиану

Delta					
RendVendable	RendPetit	RendMoy	RendGros		
14,78319259	0,462054187	-0,385436662	11,1164586	-17,15847414	MIN
7,649693988	-1,894647897	13,63135401	-1,701269142	23,313848	MAX
4,058459067	-1,81554974	5,97895811	-7,080245737	2,281450334	Ср_Знач
3,216364286	-0,437228713	0,786025678	4,419436614	0,594544699	Медиана
-17,08000232	-0,290893483	-17,15847414	-1,842464428		
17,23173938	-2,917307117	21,13689196	-0,870183451		
0,108655077	0,72703521	0,968894683	-1,154939936		
5,316129684	-3,274243878	2,068079493	7,337487539		
2,048666767	-2,724051692	0,837413991	2,44483697		
-10,59612727	-0,918194974	-4,018848089	-5,177996144		
13,87641495	-1,370726122	2,754634799	8,894619117		
21,9929436	23,313848	-1,302320789	-3,639520586		

Рис. 22 «Оценки разницы характеристик»

В данной таблице отображены:

- Delta – характеристики разности между элементами таблицы Real с соответствующими элементами таблицы Predict

- MIN – минимальное отклонение всего алгоритма относительно всей таблицы

- MAX – максимальное отклонение всего алгоритма относительно всей таблицы

- Ср_Знач – среднее значение отклонения всей таблицы

- Медиана – медиана отклонения всей таблицы

Так же вывод графиков для сравнения характеристик

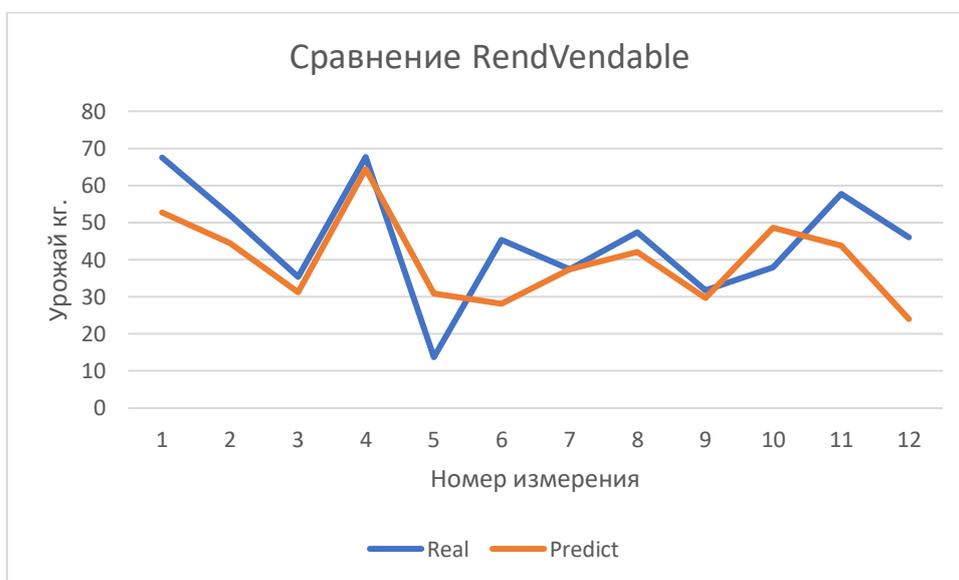


Рис. 23 «Графическое сравнение RendVendable»

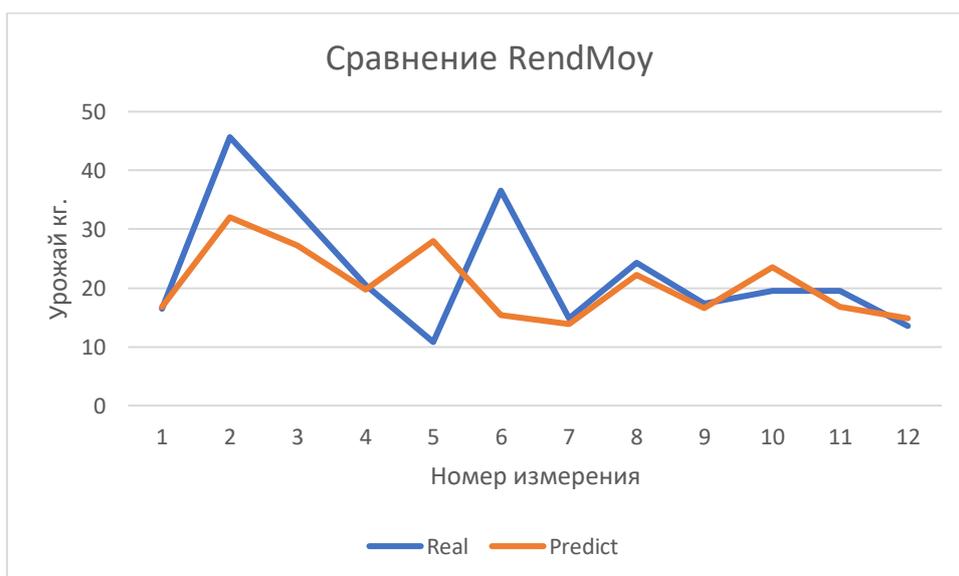


Рис. 24 «Графическое сравнение RendMoy»

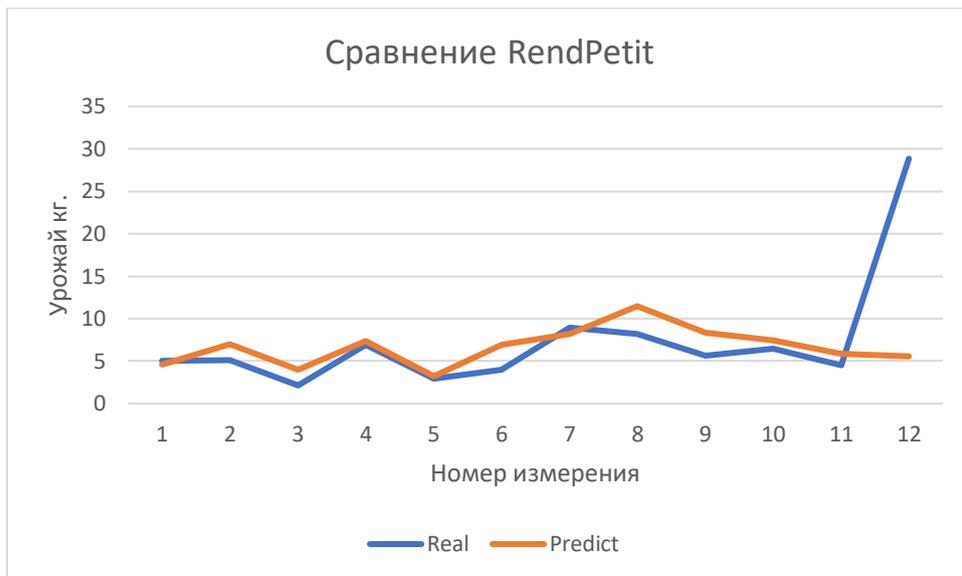


Рис. 25 «Графическое сравнение RendPetit»

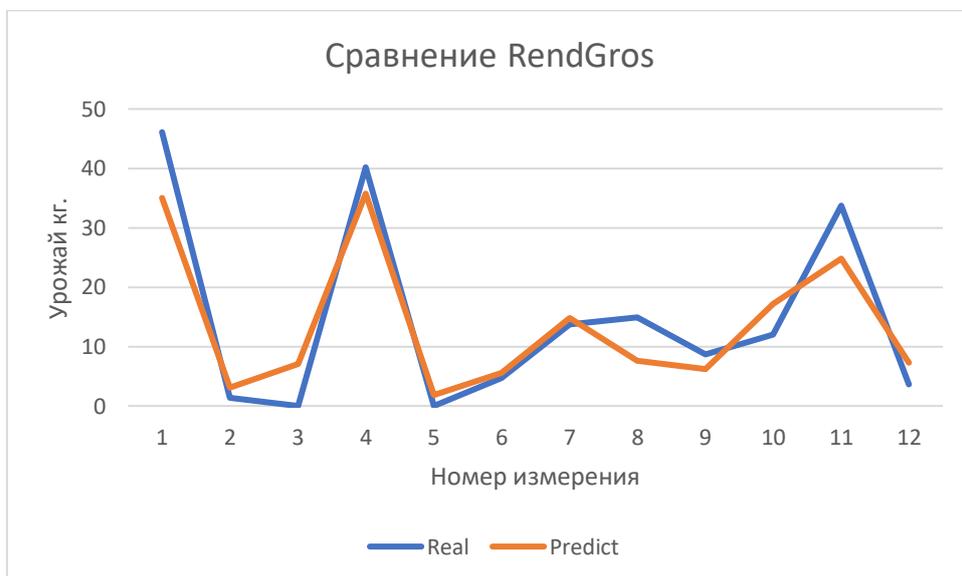


Рис. 26 «Графическое сравнение RendGros»

Как можно наблюдать, предсказательная модель достаточно хорошо показывает себя на всех типах данных, имея максимальное отклонение не более 30%.

5. Раздел «Стартап как диплом»

5.1 Описание продукта

Наименование проекта: Программное Обеспечения, выполняющее функцию мониторинга и предсказания урожайности

Краткое описание проекта:

Система подразумевает автоматический мониторинг химического состава почвы, а также определение степени влияния той или иной характеристики с целью мониторинга факторов, для определения оптимального содержания параметра и предсказания урожайности

Наименование организации: -----

Организационно-правовая форма: Индивидуальный предприниматель

Инициатор проекта: Чернюгов Кирилл Викторович

Контактные данные: тел. +7 (913) 867-42-83

Юридический адрес: -----

Фактический адрес: -----

Банковские реквизиты: -----

Краткие сведения о квалификации управленческого персонала:

- наличие технического образования по специальностям 02.03.01, 09.03.02, 09.03.03, 09.03.04 и 01.03.02;

- Знания в области садоводства, принципов ухода за растениями.

Продукт используется:

- В частном садоводческом секторе, при выращивании продукции для своих нужд

- В малом бизнесе для выращивания продукции с целью сбыта
В крупных аграрных комплексах

5.2 Интеллектуальная собственность

На момент реализации стартапа подобная технология не является

запатентованной, и при проведении анализа не может таковой являться. Ближайшая запатентованная работа, которая может иметь пересечение с нашим проектом это US20050234691A1 - Crop yield prediction. Данная работа так же анализирует ряд параметров для предсказания, но с использованием линейной регрессии, что в нашей работе не применяется (см раздел основной ВКР)

Проектная система мониторинга использует общедоступные алгоритмы анализа данных что не позволяет работе проходить по патенту как «изобретение», планируется реализация в формате «Полезная модель».

Стоит отметить, что при сложившейся экономической ситуации в мире, работа с иностранными бюро может быть проигнорирована ввиду невозможности сотрудничества, но только при условии сохранении политэкономических тенденций.

При дальнейшем развитии проекта в формат работы общего приложения планируется создание собственного торгового знака, что позволит ограничить права использования наших разработок в рамках защиты интеллектуальной собственности

Таким образом проект может представлять собой качественную самостоятельную единицу и занять некоторый объем рынка

5.3 Объем и емкость рынка. Анализ современного состояния и перспектив

История становления рынка.

Стартом данной технологии стоит отметить увеличение вычислительных мощностей, так как теперь при больших данных, программное обеспечение может гораздо проще работать с многочисленными параметрами.

Наша же технология дорабатывается существующую пересечением её с IoT (интернетом вещей), благодаря чему происходит внедрение «новой» технологии в привычные системы

Объём АПК рынка



Рис.27. доля крупнейших игроков АПК сектора

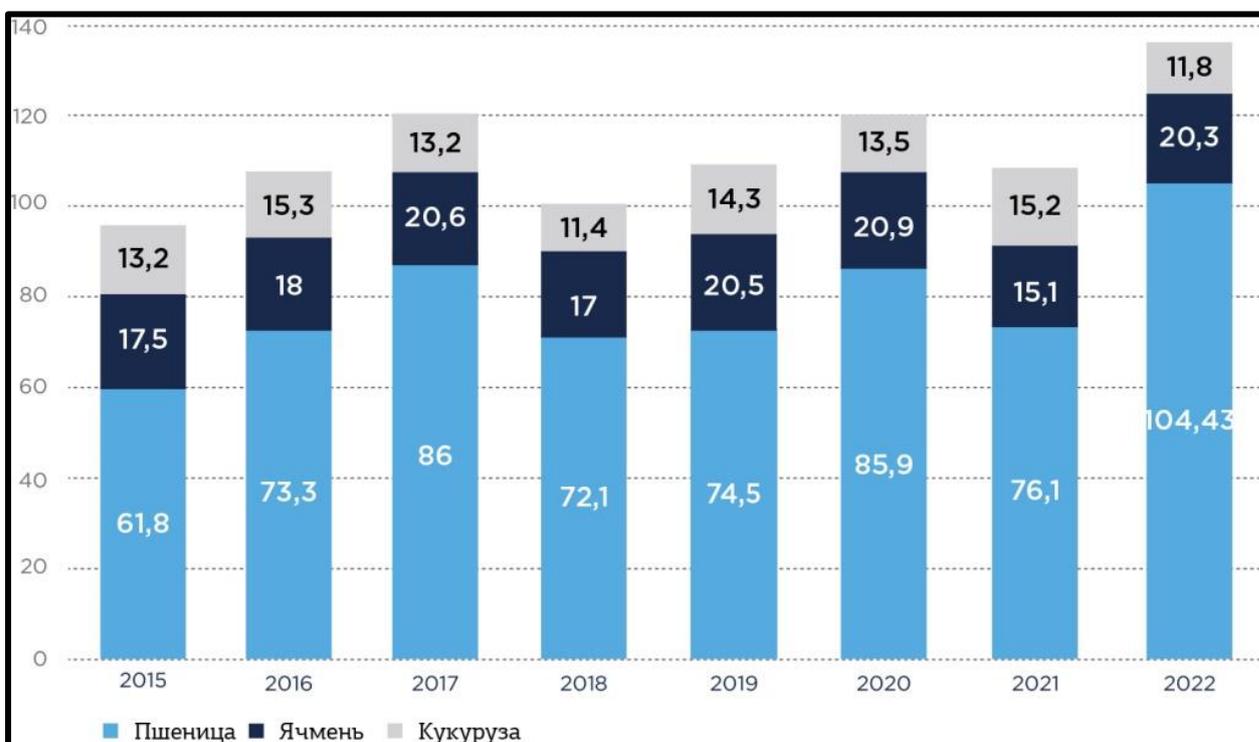


Рис.28. Валовой сбор зерновых в разрезе ряда культур, млн.т.

№	Название проекта	Расположение проекта	Объем инвестиций млн	Инвестор	Проектная мощность	Год завершения
1	Создание мясоперерабатывающего комплекса мощностью 503,5 тыс. тонн свинины в убойном весе	Курская область	96	«Мираторг»	4,5 млн голов свиней в год	2028
2	Завод по убою, хранению и переработке свинины	Тульская область	48	Группа «Черкизово»	4,2 млн голов в год	н/д
3	Строительство тепличного комплекса	Псковская область	31,94	ГК «Кабош»	50 тыс. тонн овощей и салатов в год	2024
4	Строительство высокотехнологичных теплиц общей площадью 200 га в Дербентском районе	Дагестан	30,6	АПХ «ЭКО-культура»	99,2 тыс. тонн томатов в год	2024
5	Строительство завода по глубокой переработке зерна	Красноярский край	30	«Сибагро»	н/д	2023
6	Строительство маслоэкстракционного завода и элеваторного комплекса	Курская область	28	«Содружество»	1 млн тонн сои в год	2022

7	Строительство 12 овцеводческих ферм на 1,3 млн голов	Тульская область, Курская область	27,5	«Мираторг»	35,5 тыс. тонн баранины в год	2023
8	Тепличный комбинат «Мичуринский» (II очередь)	Тамбовская область	25	Группа компаний «РОСТ»	до 100 тыс. тонн овощей и зелени в год	2023
9	Строительство сырного завода «Сибирская академия молочных наук»	Новосибирская область	24,5	«Эконива»	1,15 тыс. тонн молока в сутки	2024
10	Создание свиноводческого комплекса, строительство завода по убою и глубокой переработке мяса	Псковская область	22	Великолукский агропромышленный комплекс	400 тыс.	

Табл. 1. 10 крупнейших инвестпроектов в сфере АПК России

По имеющимся данным можно сделать вывод о достаточно больших объёмах инвестиций в сектор АПК. Нас же интересуют примеры 3,4,8 табл. 1

Эти данные позволяют нам определить объёмы заинтересованности государственных структур и частных проектов в реализации именно сельскохозяйственных проектов, в топе 10, около 30%, и один проект, однозначно нацеленный на высокотехнологичную реализацию тепличного хозяйства

В дальнейшем следует рассмотреть объёмы сельскохозяйственного комплекса отдельно

Объём сельскохозяйственного комплекса

В 2020—2022 годах потребление овощей стагнировало из-за форс-мажорных обстоятельств, но в 2023-м оно должно вырасти. Об этом в ходе организованного «Агроинвестором» форума «“Магнит” и агробизнес: партнерство в интересах покупателей» рассказала исполнительный директор Центра отраслевой экспертизы Россельхозбанка Людмила Малышева. «Из-за низкой платежеспособности в некоторых регионах люди не способны купить продукцию, которая несет большие логистические затраты. Но, если рассматривать Дальний Восток, где выращивается мало овощей, население в регионе платежеспособно», — отметила эксперт. По ее мнению, Дальневосточный федеральный округ имеет потенциал в этом направлении. Также в регионе перспективно тепличное производство высокомаржинальных продуктов: зелени, грибов и ягод. Спрос и предложение этих товаров, по ее словам, не растут по всей стране.

Технологический охват рынка сельского хозяйства



Рис.28. Структура рынка интеллектуального сельского хозяйства

Оценка объёма рынка

Период оценки: Год

Границы: Россия, с упором на области центральной Сибири и Дальнего востока

Критерии:

- 1) Уровень потребления сельскохозяйственной продукции
- 2) Производственные возможности территорий
- 3) Объемы технологической внедряемости

Аудитория и Товарные группы: Малый, Крупный бизнес и предприятия государственного сектора

Единицы измерения: Объемы производств, валюта

Источники:

- 1) <https://www.agroinvestor.ru/>
- 2) <https://www.europeanproceedings.com/article/10.15405/epsbs.2022.0>

2.21

Показатели	Годы			Отношение 2021 г. к 2019 г., %
	2019	2020	2021	
Объем произведенных инновационных товаров, работ и услуг, млн руб.	22222,9	28446,0	33829,1	152,2
Доля инновационной продукции в общем объеме отгруженных товаров и оказанных услуг, %	1,4	1,8	1,9	0,5
Затраты на технологические инновации, млн руб.	14963,3	15806,0	21960,5	146,8
Доля организаций, внедряющих технологические инновации, %	3,4	3,1	3,7	0,3

Табл. 2 «Затраты сельскохозяйственных компаний на тех. инновации»

Исходя из описанных данных, можем наблюдать рост показателя затрат на технологические инновации, что говорит о возможности входа на рынок проектной технологии

Стоит отметить, что на данном этапе проект является лишь написанием кода для системы, то есть денежная оценка может происходить сдельно без фиксированного порога цены

Уровень и тенденции развития.

На данный момент рынок находится в переходном состоянии, но ввиду

активного развития автоматизированных систем, данная область так же будет развиваться, так как их зависимость линейна.

Барьеры для входа на рынок.

- 1) Наличие устоявшихся конкурентов
- 2) Слабое освоение технологии
- 3) Слабое внедрение в обиход

Специфические особенности рынка.

- 1) Индивидуальность оборудования
- 2) Невозможность постоянного мониторинга

Факторы, благоприятствующие выходу на рынок.

- I. Всеобщая цифровизация
- II. Общее снижение умения работы с растительными культурами

5.4 Планируемая стоимость

При реализации подобных проектов в смежных средах, был выявлен числовой диапазон от 100 000, до 1 500 000 рублей, качественная оценка реализации продукта на основе стоимости его производства с оглядкой на аналогичные проекты может быть представлена в виде:

Выплаты сотрудникам	Часов работ на час разработки	Средний оклад, gros, руб	Затраты в час	Приведенные затраты в час
Проектирование (Проектировщик)	35%	40 000	401,42	140,5
Разработка (Разработчик)	100%	50 000	632,95	632,95
Тестирование (Тестировщики)	50%	30 000	427,96	213,98
Нагрузочное тестирование (Внедрение)	5%	50 000	501,78	25,09
Внедрение (Внедрение)	5%	50 000	501,78	25,09
Техническая документация (Производство)	5%	30 000	632,95	31,65
Итого выплат сотрудникам на 1 час разработки				1 069,26
Накладные общехозяйственные расходы в пересчете на час разработчика руб/час				213,85
Итоговая себестоимость				1283,11

Табл. 3 «Калькуляция стоимости часа разработки»

	Разработчик	Тестировщик	Внедрение	Проектировщики
Проектирование (Проектировщик)	50 000	30 000	50 000	40 000
Мед. Страховка (1% от зп/мес)	500	300	500	400
Компенсация питания (2% от зп/мес)	1000	600	1000	800
Ставка социальных налогов	14%	14%	14%	14%
Социальные налоги сотрудника руб/мес	7 000	4 200	7 000	5 600
Премия 15% от ЗП, руб/мес	7 500	4 500	7 500	6 000
Соц. налоги на премию руб/мес	9 868,42	5 696,2	9 868,42	7 894,74
Рабочих дней в году	247	247	247	247
Продолжительность отпуска, раб. дней	20	20	20	20
Рабочих дней в месяце с учетом отпуска	18,9	18,9	18,9	18,9
Затраты времени в день на сопровождение кода и поддержку в актуальном состоянии, час	2,4	2,4		
Затраты времени на разработку, час	5,6	5,6		
Затраты времени в день, час			8,0	8,0
Расходы на оплату труда в перерасчете на час разработки, руб	623,68	427,96	501,78	401,42

Табл. 4 «Калькуляция затрат на оплату труда часа работы сотрудника»

Сроки реализации проекта начинаются от месяца, при перерасчете на себестоимость, стоимость проекта выходит от 200 000, таким образом наш продукт входит в рамки начального проекта и ценового диапазона от 100 000 до 500 000

Таким образом на данный момент цена может являться нашим конкурентным преимуществом

5.5 Конкурентные преимущества

Продукция конкурентов:

Agrivi – ПО для анализа состояния урожая, поиска вредителей и календарь работ

Granular - Отчетность полей для офиса

Trimble – Мониторинг отчетов сельскохозяйственной деятельности

FarmERP – Финансовая составляющая фермерской работы

Спрос на продукцию конкурентов.

Основной спрос возникает у крупных игроков рынка, ввиду работы с большими объемами информации с больших объемов полей

Ценообразование.

Главный фактором ценообразования продукции конкурентов является их технологические решения и реализация плана по подписке

Анализ конкурентов

Конкуренты\ Критерии	Agrivi	Granular	Trimble	FarmERP	Мы
Качество	4	4	5	3	3
Цена	2	3	4	5	5
Мобильность	4	3	3	3	3
Инновационность	4	2	4	2	5
Итого	19	14	15	15	19

Табл. 5 «Оценка конкурентов»

Как можно наблюдать, продукция конкурентов может хоть и по уровню удобства превосходить наш продукт, но ввиду инновационности разработки, система мониторинга может вполне конкурировать не только на местном рынке, но и на мировом.

5.6 Целевые сегменты потребителей

Для выделения потребителя, необходимо выделить занимаемый им сегмент на рынке, а также необходимо выделить сам рынок, на который ориентирует проект

Рынок: Сельскохозяйственная продукция

Сегмент: Компании, занимающиеся растительным производством

Примеры компаний:

- 1) ГК «Содружество»
- 2) «Русагро»
- 3) «Мираторг»
- 4) ГАП «Ресурс»
- 5) ГК «Кабош»
- 6) Группа компаний «РОСТ»
- 7) АПХ «ЭКО-культура»

Примеры заказов смежных с нашим проектом:

- 1) Тепличный комбинат «Мичуринский» (II очередь) - Группа компаний «РОСТ» (25 млн руб инвестиций)
- 2) Строительство высокотехнологичных теплиц общей площадью 200 га в Дербентском районе - АПХ «ЭКО-культура» (30,6 млн руб инвестиций)
- 3) Строительство тепличного комплекса - ГК «Кабош» (31,94 млн руб инвестиций)

Критерии привлекательности:

- 1) Большие объемы инвестиций (от 10 млн руб)
- 2) Слабая конкурентная среда из-за спецификации проекта
- 3) Геополитическая нестабильность, из-за чего требования к импортозамещению. Продукция подобного толка на данный момент исключительно импортная
- 4) Сегментация рынка
- 5) Возможность переход в смежные области ввиду разветвленности структур распределения (добавление к будущему приложению не только анализа производств, но также и ПО бух учета и экономического анализа, логистического анализа)

Итоговая оценка:

Данный рынок привлекателен в первую очередь своими размерами, низкой конкуренцией и высоким спросом на реализацию отечественных решений. Отсюда принято решение направить проект по данному сегменту рынка

5.7 Бизнес модель проекта

На данный момент проект может существовать исключительно в сделальном формате, то есть ценообразование идет по договоренности с нанимателем

В дальнейшем возможна интеграция в системы на уровне приложения, что позволит перейти проекту на подписочный уровень, и обеспечит компании постоянную прибыль

Бизнес модель по Остервальдеру

<p>Ключевые партнёры</p> <ul style="list-style-type: none"> • Компании занимающиеся непосредственно химическим анализом почвы • Компании разрабатывающие крупные системы для АПК 	<p>Ключевые действия</p> <ul style="list-style-type: none"> • Проведение статистического анализа • Развитие собственного приложения 	<p>Ценностное предложение</p> <p>Проведение статистического анализа с целью оценки прогноза урожая</p>	<p>Взаимоотношение с клиентами</p> <p>Рекомендации от искусственого интеллекта + собственный сервис</p>	<p>Сегменты клиентов</p> <ul style="list-style-type: none"> • Сельскохозяйственные производители <ul style="list-style-type: none"> ○ АПК комплекс
<p>Ключевые ресурсы</p> <p>Товарный знак Набор датчов результата обучения модели</p>		<p>Каналы</p> <ul style="list-style-type: none"> • Гос. Контракты • Частные <u>контракты с фирмами</u> описанными при анализе рынка 		
<p>Структура издержек</p> <ul style="list-style-type: none"> • Разработка ПО • Разработка и развитие приложения • Развитие технологической платформы 		<p>Потоки доходов</p> <ul style="list-style-type: none"> • Работа на заказ • При работе приложения - подписка 		

Табл. 6 «Бизнес модель по Остервальдеру»

Вид деятельности

Код ОКВЭД	Деятельность
62.01	Разработка компьютерного программного обеспечения
58.29	Издание прочих программных продуктов
62.02	Деятельность консультативная и работы в области компьютерных технологий
58.2	Издание программного обеспечения
63.11.1	Деятельность по созданию и использованию баз данных и информационных ресурсов
62.03.12	Деятельность по управлению компьютерными системами дистанционно
47.78.9	Торговля розничная непродовольственными товарами, не включенными в другие группировки, в специализированных магазинах
47.76.1	Торговля розничная цветами и другими растениями, семенами и удобрениями в специализированных магазинах
47.41.2	Торговля розничная программным обеспечением в специализированных магазинах
47.91.1	Торговля розничная по почте
47.91.2	Торговля розничная, осуществляемая непосредственно при помощи информационно-коммуникационной сети Интернет
96.09	Предоставление прочих персональных услуг, не включенных в другие группировки

5.8 Итог

При анализе финансовых возможностей реализации данного проекта, был сделан вывод, что на данном этапе возможна только реализация в рамках заказной разработки программного обеспечения. В дальнейшем продукт, при увеличении его технологических возможностей может перейти в новую форму, в формат приложения, с постоянной подпиской и новыми функциями.

6. Раздел «Социальная ответственность»

6.1 Введение

Охрана труда и окружающей среды, а также обеспечения безопасности рабочих в чрезвычайных ситуациях являются важными компонентами организации рабочего процесса в настоящее время. В современной жизни компьютер широко применяется в жизни человека: и дома, и в офисе, и в магазине, и в производстве, и даже в бытовой технике. Другими словами, компьютеры прочно вошли в повседневную жизнь людей и их использование постоянно увеличивается. Несоблюдение требований безопасности приводит к тому, что при работе за компьютером сотрудник может ощущать дискомфорт: возникают головные боли и резь в глазах, появляются усталость и раздражительность. У некоторых людей нарушается сон, аппетит, ухудшается зрение, начинают болеть руки, шея, поясница и тому подобное.

При ненормированной работе возможно нервное истощение.

Внедрение финорганизации и охране труда.

В данной выпускной квалификационной работе проводится оценка урожайности картофеля с опорой на химический состав почвы с помощью 3 широко известных модели, затем выбирается модель, которая больше всего подходит для данной задачи.

В настоящее время в производстве, научно-исследовательских и конструкторских работах, сфере управления и образования персональные ЭВМ (ПЭВМ) находят все большее применение. Компьютеры уже завоевали свое место на предприятиях, в организациях, офисах и даже в домашних условиях.

Однако компьютер является источником вредного воздействия на организм человека, а, следовательно, и источником профессиональных заболеваний. Это предъявляет к каждому пользователю персонального

компьютера требование – знать о вредном воздействии ПЭВМ на организм человека и необходимых мерах защиты от этих воздействий.

Предметом исследования является рабочая зона студента, включающая письменный стол, персональный компьютер, клавиатуру, компьютерную мышь, стул, а также само помещение, в котором находится рабочая зона.

Характеристики помещения:

- минимальная площадь помещения – 14 м²;
- высота помещения – 3 м;
- число и размер окон – 1 (1,3×1,4);
- вентиляция – приточно-вытяжная;
- число рабочих мест – 1.

Основной целью раздела социальной ответственности является создание оптимальных норм для обеспечения производственной безопасности человека и создания комфортных условий труда.

6.2 Правовые и организационные вопросы обеспечения безопасности

Согласно Трудовому кодексу РФ, работодатель должен 2 раза в месяц выплачивать заработную плату, делать выплаты в Пенсионный фонд РФ, обеспечивать нормальную продолжительность рабочего времени, оплачиваемые отпуска и режим отдыха на рабочем месте. Организация должна предоставлять ежегодные отпуска продолжительностью 28 календарных дней.

Трудовой кодекс определяет нормальную продолжительность трудового времени максимально 40 часов в неделю. Под рабочим временем подразумевается время, в течение которого сотрудник должен выполнять трудовые обязанности в соответствии с трудовым договором.

Основная работа происходит за компьютерным устройством, что соответствует группе III и подразумевает до 6 часов работы за компьютером.

Продолжительность же непрерывной работы за компьютером не должна превышать 2-х часов с регламентированными перерывами по 20 минут. Основным нормативным документом, определяющим требования к организации рабочих мест пользователей ПЭВМ, является ГОСТ 12.2.032-78 ССБТ «Рабочее место при выполнении работ сидя. Общие эргономические требования», с общими положениями:

1) Рабочее место для выполнения работ сидя организуют при легкой работе, не требующей свободного передвижения работающего, а также при работе средней тяжести в случаях, обусловленных особенностями технологического процесса. Категории работ - по ГОСТ 12.1.005-88.

2) Конструкция рабочего места и взаимное расположение всех его элементов (сиденье, органы управления, средства отображения информации и т.д.) должны соответствовать антропометрическим, физиологическим и психологическим требованиям, а также характеру работы.

3) Рабочее место должно быть организовано в соответствии с требованиями стандартов, технических условий и (или) методических указаний по безопасности труда.

В соответствии с СанПиНом 1.2.3685-21 «Гигиенические требования к персональным электронно-вычислительным машинам и организации работы» рабочие места с ПК по отношению к световым проемам должны располагаться так, чтобы естественный свет падал сбоку, желательно слева. Схемы размещения рабочих мест с ПК должны учитывать расстояние между рабочими столами с мониторами: расстояние между боковыми поверхностями мониторов не менее 1,2 м, а расстояние между экраном монитора и тыльной частью другого монитора не менее 2 м.

В соответствии с ГОСТ 12.2.032-78. «Система стандартов безопасности труда». Рабочее место при выполнении работ сидя» рабочий стол может быть любой конструкции, отвечающей современным требованиям эргономики и позволяющей удобно разместить на рабочей поверхности оборудование с учетом его количества, размеров и характера выполняемой работы.

Выполнение требований на данном рабочем месте отражено ниже в таблице 7, согласно СанПиН 1.2.3685-21 и ГОСТ 12.2.032-78.

Требование	Требуемое значение	Значение параметров в помещении
Высота рабочей поверхности стола	Регулируемая высота(680-800мм) Нерегулируема высота (725мм)	Нерегулируемая высота (700мм)
Рабочий стул	Подъемно-поворотный, регулируемый по высоте и углу наклона спинки	Соответствует
Расположение монитора от глаз пользователя	600-700мм	Соответствует

Таблица 7 – Требования к организации рабочего места при работе с ПЭВМ

6.3 Производственная безопасность

Работа, которая заключается в оценке финансовой устойчивости предприятий строительной отрасли России с помощью 6 широко известных моделей проводится на персональном компьютере (ПЭВМ).

При выполнении работ на персональном компьютере (ПЭВМ) согласно «ГОСТ 12.0.003-2015 Система стандартов безопасности труда (ССБТ). Опасные и вредные производственные факторы. Классификация» могут иметь место следующие факторы, представленные в таблице 8.

Факторы (ГОСТ 12.0.003-2015)	Нормативные документы
1. Отклонение показателей микроклимата в закрытом помещении	СанПиН 1.2.3685-21 Гигиенические требования к микроклимату производственных помещений.
2. Недостаточная освещенность	СП 52.13330.2016 «Естественное и искусственное освещение». Актуализированная редакция СНиП 23-05-95
3. Нервно-психические перегрузки, монотонность трудового процесса	Трудовой кодекс Российской Федерации от 30.12.2011 N 197-ФЗ (ред. от 09.03.2021);
4. Повышенное значение напряжения в электрической цепи, замыкание которой может произойти через тело человека	ГОСТ 12.1.038-82 ССБТ. Электробезопасность. Предельно допустимые уровни напряжений прикосновения и токов.

Таблица 8 – Возможные опасные и вредные факторы

1) Отклонение показателей микроклимата в закрытом помещении

Микроклимат определяется действующими на организм человека показателями температуры, влажности и скорости движения воздуха. Длительное воздействие на человека неблагоприятных показателей микроклимата ухудшает его самочувствие, снижает производительность труда и приводит к заболеваниям, поэтому в организации должны обеспечиваться оптимальные параметры микроклимата, установленные СанПиН 1.2.3685-21 «Гигиенические требования к микроклимату производственных помещений». Они представлены в таблице 9:

Период года	Категория работ по уровню энергопотребления, Вт	Температура воздуха, °С		Температура воздуха по поверхности	Относительная влажность воздуха	Скорость движения воздуха	
		Диапазон ниже оптимальных величин	Диапазон выше оптимальных величин			Для диапазона температур воздуха ниже оптимальных величин, не более	Для диапазона температур воздуха выше оптимальных величин, не более
Холодный	Ia (до 139)	20,0 - 21,9	24,1 - 25,0	19,0 - 26,0	15 - 75	0,1	0,1
Тёплый	Ia (до 139)	21,0 - 22,9	25,1 - 28,0	20,0 - 29,0	15 - 75	0,1	0,2

Таблица 9 - Допустимые нормы микроклимата в рабочей зоне производственных помещений

2) Недостаточная освещенность

Недостаточная освещенность рабочей зоны помещения, оборудованной ПК, также является одной из причин нарушения зрительной функции, а также влияет на общее самочувствие и эффективность труда. Искусственное освещение в помещениях для эксплуатации ПК должно осуществляться системой общего равномерного освещения. В случаях преимущественной работы с документами, следует применять системы комбинированного освещения (к общему освещению дополнительно устанавливаются светильники местного освещения, предназначенные для освещения зоны расположения документов). Окна в помещениях, где эксплуатируется вычислительная техника, преимущественно должны быть ориентированы на север и северо-восток. Нормируемые показатели естественного, искусственного и совмещенного освещения в соответствии с СП 52.13330.2016 указаны в таблице 10.

	Рабочая Поверхность и плоскость нормирования	Естественное освещение		Совмещенное освещение	
		КЕО е _н , %		КЕО е _н , %	
Помещения	КЕО и освещенности (Г-горизонтальная, В-вертикальная) и высота плоскости над полом, м	При верхнем или комбинированном освещении	При боковом освещении	При верхнем или комбинированном освещении	При боковом освещении
Кабинеты, рабочие комнаты, офисы, представительства	Г – 0,8	3,0	1,0	1,8	0,6
Помещения	Искусственное освещение				
	Освещенность, лк				
	При комбинированном освещении	При общем освещении	Показатель дискомфорта, М, не более	Коэффициент пульсации освещенности, %, не более	К _п
Всего	От общего				
Кабинеты, рабочие комнаты, офисы, представительства	400	200	300	40	15

Таблица 10- Нормируемые показатели естественного, искусственного и Совмещенного освещения

3) Нервно-психические перегрузки

Нервно-психические перегрузки – совокупность таких сдвигов в психофизиологическом состоянии организма человека, которые развиваются после совершения работы и приводят к временному снижению эффективности труда. Состояние утомления (усталость) характеризуется определенными объективными показателями и субъективными ощущениями.

Нервно-психические перегрузки подразделяются на следующие:

- умственное перенапряжение;
- перенапряжение анализаторов;
- монотонность труда;

- эмоциональные перегрузки.

При первых симптомах психического перенапряжения необходимо:

- дать нервной системе расслабиться;
- рационально чередовать периоды отдыха и работы;
- начать заниматься спортом; - ложиться спать в одно и то же время;
- в тяжелых случаях обратиться к врачу.

Естественно, что полностью исключить провоцирующие факторы из жизни вряд ли удастся, но можно уменьшить их негативное воздействие, давая нервной системе необходимый отдых.

4) Повышенное значение напряжения в электрической цепи, замыкание которой может произойти через тело человека.

В деятельности организации широко используется электричество для питания компьютерной техники, которая может являться источником опасности. Несоблюдение правил ГОСТ 12.1.038-82 ССБТ. «Электробезопасность. Предельно допустимые уровни напряжений прикосновения и токов» может привести к опасным последствиям. Поражение электрическим током может произойти при прикосновении к токоведущим частям, находящимся под напряжением, на которых остался заряд или появилось напряжение.

Электрический ток оказывает на человека термическое, электролитическое, биологическое и механическое воздействие. Действие электрического тока на человека приводит к травмам или гибели людей. Для переменного тока частотой 50 Гц допустимое значение напряжения прикосновения составляет 2 В, а силы тока – 0,3 мА, для тока частотой 400 Гц, соответственно – 2 В и 0,4 мА, для постоянного тока – 8 В и 1 мА. Мерами защиты от воздействия электрического тока являются оградительные устройства, устройства автоматического контроля и сигнализации, изолирующие устройства и покрытия, устройства защитного заземления, устройства автоматического отключения, предохранительные устройства.

6.4 Экологическая безопасность

Объект исследования является теоретическим и не оказывает влияния на окружающую среду

В ходе данной работы были использованы следующие ресурсы:

- электроэнергия для работы компьютера;
- бумага;
- люминесцентные лампы.

С точки зрения потребления ресурсов компьютер потребляет сравнительно небольшое количество электроэнергии, что положительным образом сказывается на общей экономии потребления электроэнергии в целом.

При написании ВКР вредных выбросов в атмосферу, почву и водные источники не производилось, радиационного заражения не произошло, чрезвычайные ситуации не наблюдались, поэтому не оказывались существенные воздействия на окружающую среду, и никакого ущерба окружающей среде не было нанесено.

В связи с тем, что огромная масса информации содержится на бумажных носителях, уничтожение бумаги играет очень важную роль. Среди основных методов уничтожения, которые применяются на сегодняшний день для бумажных документов, следует отметить следующие:

- Сжигание документов.
- Шредирование.
- Закапывание.
- Химическая обработка.

Вышедшее из строя ПЭВМ и сопутствующая оргтехника относится к IV классу опасности и подлежит специальной утилизации. Для оказания наименьшего влияния на окружающую среду, необходимо проводить

специальную процедуру утилизации ПЭВМ и оргтехники, при которой более 90% отправится на вторичную переработку и менее 10% будут отправлены на свалки. При этом она должна соответствовать процедуре утилизации ГОСТ Р53692-2009 Ресурсосбережение. Обращение с отходами. Этапы технологического цикла отходов.

В ходе деятельности организация также создает бытовой мусор (канцелярские, пищевые отходы, искусственные источники освещения), который должен быть утилизирован в соответствии с определенным классом опасности или переработан, чтобы не оказывать негативное влияние на состояние литосферы.

Ртуть из разбитых люминесцентных ламп под влиянием бактерий на полигонах ТБО преобразуется в токсичную водорастворимую метилртуть. Она накапливается в грунтах, открытых водоемах и подземных источниках, а затем – в организмах рыб и моллюсков.

Метилртуть, которую человек получает из пищи рыбного происхождения, отравляет нервную систему, наносит непоправимый вред пищеварению и иммунитету, легким и почкам. Если ее концентрация зашкалит, возможен и летальный исход. При умеренной интоксикации наблюдаются постоянные головные боли, потеря памяти, бессонница, нарушения моторной и когнитивной функции.

Согласно требованиям санитарных норм СанПиН 2.1.3684-21, в медучреждениях и других организациях неисправные люминесцентные лампы хранят отдельно от других отходов в специальном закрытом помещении, защищенном от осадков, грунтовых и поверхностных вод. Главное условие – обеспечить целостность осветительных приборов, поэтому они должны быть в жесткой таре, например в металлическом контейнере с герметичной крышкой и чехлом. Лампы укладывают плотно, чтобы они не разбились во время транспортировки.

Согласно правилам утилизации ртутьсодержащих отходов: Люминесцентные лампы упаковывают в специальные маркированные

контейнеры с чехлом и герметичными крышками, градусники – в емкости, напоминающие по форме термос (тара не должна быть желтой или красной).

Перевозить отработанные люминесцентные источники освещения можно в специальных контейнерах, которые уберегут стеклянные корпуса от повреждений. Существует несколько вариантов утилизации:

- Демеркуризация. Лампы подают на конвейере в специальную дробильную установку. Пары ртути осаждают сорбентом.

- Термокриогенная утилизация. Ртуть попадает в глубоковакуумную ловушку при температуре 170 °С. Вредные пары замораживают жидким азотом. По окончании ртуть размораживается и стекает в приемник.

- Дробление. Под давлением воздуха в противоточной системе лампа распадается в разные приемники на стеклобой, цоколь и люминофор. Этот метод – наиболее бюджетный, но уступает другим по эффективности.

6.5 Безопасность в чрезвычайных ситуациях

С учетом специфики работы и наличием вычислительной техники в помещении наиболее вероятно возникновение пожара, под которым понимается вышедший из-под контроля процесс горения, обусловленный возгоранием вычислительной техники и угрожающий жизни и здоровью работников.

Причинами возгорания при работе с компьютером могут быть:

- токи короткого замыкания;
- неисправность устройства компьютера или электросетей;
- небрежность оператора при работе с компьютером;
- возпламенение ПК из-за перегрузки.

В связи с этим, согласно ГОСТ 12.1.004-91 «Пожарная безопасность. Общие требования», при работе с компьютером необходимо соблюдать следующие нормы пожарной безопасности:

- для предохранения сети от перегрузок запрещается одновременно подключать к сети количество потребителей, превышающих допустимую нагрузку;

–работы за компьютером проводить только при исправном состоянии оборудования, электропроводки; – иметь средства для тушения пожара (огнетушитель);

–установить количество, размеры и соответствующее конструктивное исполнение эвакуационных путей и выходов;

–обеспечить возможность беспрепятственного движения людей по эвакуационным путям.

–Прокладка всех видов кабелей в металлических газонаполненных трубах – отличный вариант для предотвращения возгорания. При появлении пожара, любой, увидевший пожар должен: незамедлительно заявить о данном факте в пожарную службу по телефонному номеру 01 или 112, заявить о происшествии и соблюдать покой.

6.6 Выводы по разделу

Проанализировав и оценив нормы условий труда в рабочем помещении, где была разработана выпускная квалификационная работа, можно сделать выводы, о том что помещение должно удовлетворять всем нормативным требованиям. Действие вредных и опасных факторов должно быть сведено к минимуму, т.е. микроклимат, освещение и электробезопасность соответствовать требованиям, предъявленным в соответствующих представленных нормативных документах.

- Категория помещения по электробезопасности: Безопасные помещения

- Группа персонала по электробезопасности согласно Правил по охране труда при эксплуатации электроустановок: Группа 1

- Категория тяжести труда по СанПиН 1.2.3685-21 "Гигиенические нормативы и требования к обеспечению безопасности и (или) безвредности для человека факторов среды обитания": 1б с энергозатратами 140-174 Вт

- Помещение по взрывопожарной и пожарной опасности согласно СП 12.13130.2009 «Определение категорий помещений, зданий и наружных

установок по взрывопожарной и пожарной опасности»: Категория Г

- Так же указывается категория объекта, оказывающего значительное негативное воздействие на окружающую среду: Категория IV

Не стоит забывать, что монитор компьютера служит источником вредного фактора и отрицательно влияет на здоровье офисного сотрудника. Во избежание этого, нужно делать перерывы в работе и проводить специальные комплексы упражнений для разминки тела.

Заключение

В данной работе проведен подробный анализ набора данных. Продемонстрирована целесообразность его применения в климатических зонах России. Проанализированы три регрессионные модели - случайные деревья, линейная регрессия и многослойный перцептрон, и исходя из оценок их эффективности выбрана наиболее точная – RandomForest со средними показателями:

- R^2 (Коэффициент детерминации) – 0,67;
- Средняя абсолютная ошибка в процентах (MAPE) – 0,35
- средняя абсолютная ошибка (MAE) – 4
- средняя квадратичная ошибка (RMSE) – 5
- Абсолютная ошибка – 16

Оформлен конечный вывод данных в таблицу, проведены сравнения данных предсказанных с реальными.

При анализе проделанной работы модель показала свою работоспособность, а также достаточно приемлемую точность, благодаря чему её можно использовать как алгоритм построения оценки урожайности картофеля с опорой на химический состав почвы и характеристики сорта.

Список литературы

1. J. M. McKinion, H. E. Lemmon. - "Expert systems for agriculture," Computers and Electronics in Agriculture, vol. 1 no. 1, pp. 31-40, 1985.
2. A G. Boulanger, —The expert system PLANT/CD: A case study in applying the general-purpose inference system ADVISE to predicting black cutworm damage in corn, || Ph.D. Thesis, University of Illinois at Urbana-Champaign, 1983.
3. J. Roach, R. Virkar, C. Drake, M. Weaver, "An expert system for helping apple growers," Computers and electronics in agriculture, vol. 2 no. 2, pp. 97-108, 1987.
4. N. D. Stone, T. W. Toman, "A dynamically linked expert-database system for decision support in Texas cotton production," Computers and electronics in agriculture, vol. 4 no. 2, pp. 139-148, 1989.
5. C. Robinson, N. Mort, "A neural network system for the protection of citrus crops from frost damage." Computers and Electronics in Agriculture, vol. 16 no. 3, pp. 177-187, 1997
6. S. K. Li, X. M. Suo, Z. Y. Bai, Z. L. Qi, X. H. Liu, S. J. Gao, S. N. Zhao, "The machine recognition for population feature of wheat images based on BP neural network," Agricultural Sciences in China, vol.1 no. 8, pp. 885-889, 2002.
7. C. Prakash, A. S. Rathor, G. S. M. Thakur, "Fuzzy based Agriculture expert system for Soyabean." in Proc. International Conference on Computing Sciences WILKES100-ICCS2013, Jalandhar, Punjab, India. 2013.
8. D. W. Boyd, M. K. Sun, "Prototyping an expert system for diagnosis of potato diseases," Computers and Electronics in Agriculture, vol. 10 no. 3, pp. 259-267, 1994.
9. S. K. Sarma, K. R. Singh, A. Singh, "An Expert System for diagnosis of diseases in Rice Plant." International Journal of Artificial Intelligence, vol.1 no.1, pp. 26-31, 2010.
10. K. Balledda, et al., "Agpest: An efficient rule-based expert system to prevent pest diseases of rice & wheat crops," in Proc. Intelligent Systems and Control (ISCO)-2014, IEEE, 2014.
11. V. Tilva, J. Patel, C. Bhatt, "Weather based plant diseases forecasting using fuzzy logic," In proc. (NUiCONE), 2013. IEEE, 2013.
12. L. J. Francl, and S. Panigrahi, "Artificial neural network models of wheat leaf wetness," Agricultural and Forest Meteorology, vol. 88 no. 1, pp. 57-65, 1997.
13. M. S. P. Babu, B. S. Rao, "Leaves recognition using back propagation neural network-advice for pest and disease control on crops." IndiaKisan. Net: Expert Advisory System (2007).
14. M. Ismail and Mustikasari, —Intelligent system for tea leaf disease detection,|| IPSJ Technical report, pp. 1-4, 2013.
15. B. C. Karmokar, et al., "Tea leaf diseases recognition using neural network ensemble," International Journal of Computer Applications, vol. 114 no.17, pp. 27-30, 2015.
16. S. Sladojevic, et al. "Deep neural networks-based recognition of plant diseases by leaf image classification," Computational intelligence and neuroscience, 2016.
17. A. M. G. J. Hanson, A. Joy, J. Francis, —Plant LeafDisease Detection using Deep Learning and Convolutional Neural Network,|| International Journal of Engineering Science, vol. 7 no. 3, pp. 5324-5328, 2017.
18. F. Hahn, I. Lopez, G. Hernandez, "Spectral detection and neural network discrimination of Rhizopus stolonifer spores on red tomatoes," Biosystems Engineering, vol. 89 no. 1, pp. 93-99, 2004.

19. K. Y. Huang "Application of artificial neural network for detecting Phalaenopsis seedling diseases using color and texture features," *Computers and Electronics in agriculture*, vol.57 no. 1, pp. 3-11, 2007.
20. S. S. Sannakki, et al. "Leaf disease grading by machine vision and fuzzy logic," *Int. J. of Comp. Tech. and Application*, vol. 2 no. 5, pp. 1709-1716, 2011.
21. H. Al-Hiary, et al., "Fast and accurate detection and classification of plant diseases," *Int. J. of Computer Application*, vol. 17 no. 1, pp. 31-38, 2011.
22. D. Al Bashish, M. Braik, S. Bani-Ahmad, "Detection and classification of leaf diseases using K-means-based segmentation and Neuralnetworks-based classification," *Information Technology Journal*, vol. 10 no.2, pp. 267-275, 2011
23. S. F. Khan, et al., "Dr. Wheat: A Web-based expert system for diagnosis of diseases and pests in Pakistani wheat." In *Proc. of the World Congress on Engineering*. Vol. 1. 2008.
24. I. A. Martynenko, and Simon X. Yang, "Biologically inspired neural computation for ginseng drying rate," *Biosystems engineering*, vol. 95 no.3, pp. 385-396, 2006.
25. K. Movagharnejad, and Maryam Nikzad, "Modeling of tomato drying using artificial neural network," *Computers and electronics in agriculture*, vol. 59 no.1, pp. 78-85, 2007.
26. P. Boniecki, Krzysztof Nowakowski, and R. Tomczak, "Neural networks type MLP in the process of identification chosen varieties of maize," in *Proc. Third International Conference on Digital Image Processing (ICDIP 2011)*, Vol. 8009. International Society for Optics and Photonics, 2011.
27. G. M. Pasqual, "Development of an expert system for the identification and control of weeds in wheat, triticale, barley and oat crops," *Computers and electronics in agriculture*, vol. 10 no. 2, pp. 117-134, 1994.
28. T. F. Burks, et al., "Backpropagation neural network design and evaluation for classifying weed species using color image texture," *Transactions of the ASAE*, vol. 43 no. 4, pp. 1029-1037, 2000.
29. T. F. Burks, et al., "Evaluation of neural-network classifiers for weed species discrimination," *Biosystems Engineering*, vol. 91 no.3, pp. 293-304, 2005.
30. G. Liu, Xuehong Yang, and Minzan Li, "An artificial neural network model for crop yield responding to soil parameters," in *Proc. International Symposium on Neural Networks*, Springer, Berlin, Heidelberg, 2005
31. D. L. Ehret, et al., "Neural network modeling of greenhouse tomato yield, growth and water use from automated crop monitoring data," *Computers and electronics in agriculture*, vol. 79 no.1, pp. 82-89, 2011.
32. K. Thongboonnak, and Sunya Sarapirome, "Integration of Artificial Neural Network and Geographic Information System for Agricultural Yield Prediction," *Suranaree Journal of Science & Technology*, vol. 18 no.1, pp. 71-80, 2011.
33. R. Pahlavan, Mahmoud Omid, and Asadollah Akram, "Energy input–output analysis and application of artificial neural networks for predicting greenhouse basil production," *Energy*, vol. 37 no. 1, pp. 171-176, 2012.
34. B. Khoshnevisan, et al., "Modeling of energy consumption and GHG (greenhouse gas) emissions in wheat production in Esfahan province of Iran using artificial neural networks," *Energy*, vol. 52, pp. 333-338, 2013.

35. A. Nabavi-Pelesaraei, et al., "Neural network modeling of energy use and greenhouse gas emissions of watermelon production systems," *Journal of the Saudi Society of Agricultural Sciences*, vol. 15 no.1, pp. 38-47, 2016.
36. F. Soheili-Fard, and Seyed Babak Salvatian, "Forecasting of tea yield based on energy inputs using artificial neural networks (a case study: Guilan province of Iran)," *Biological Forum*, Vol. 7. No. 1, pp. 1432-1438, 2015.
37. S. S. Dahikar, and Sandeep V. Rode. "Agricultural crop yield prediction using artificial neural network approach," *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering*, vol. 2 no.1, pp. 683-686, 2014
38. I. Ghosh and R. K. Samanta, "TEAPEST: An expert system for insect pest management in tea," *Applied Engineering in Agriculture*, vol. 19 no. 5, pp. 619, 2003
39. C. Arif, M. Mizoguchi, and B. I. Setiawan, —Estimation of soil moisture in paddy field using Artificial Neural Networks, arXiv preprint arXiv: 1303.1868, 2013
40. V. F. Bralts, et al., "An expert system for the hydraulic analysis of microirrigation systems," *Computers and electronics in agriculture*, vol. 9 no. 4, pp. 275-287, 1993.
41. M. Valdés-Vela, et al., "Soft computing applied to stem water potential estimation: A fuzzy rule-based approach," *Computers and Electronics in Agriculture*, vol. 115, pp. 150-160, 2015
42. ГОСТ 12.2.032-78 Рабочее место при выполнении работ сидя
43. Трудовой кодекс РФ от 30 декабря 2001 г. N 197-ФЗ (ТК РФ)
44. Белов, Сергей Викторович. Безопасность жизнедеятельности и защита окружающей среды (техносферная безопасность) : учебник для академического бакалавриата / С. В. Белов. - 5-е изд., перераб. и доп.. -Москва: Юрайт ИД Юрайт, 2015. - 703 с.
45. О.Б. Назаренко, Ю.А. Амелькович, БЕЗОПАСНОСТЬ ЖИЗНЕДЕЯТЕЛЬНОСТИ, Издательство Томского политехнического университета, 2013 г., 178 стр
46. Ларионов Н. М. Промышленная экология [Электронный ресурс] : учебник для бакалавров / Н. М. Ларионов. – Мультимедиа ресурсы (10 директорий; 100 файлов; 740МВ). – Москва : Юрайт, 2013.
47. Березин И. С. Маркетинговый анализ. Рынок. Фирма. Товар. Продвижение. 3-е изд., испр. и доп. М.: Вершина, 2008. 480 с.
48. Alex Osterwalder, Yves Pigneur, *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers*, John Wiley and Sons; First Edition (July 13, 2010), 288 pages
49. Право интеллектуальной собственности. Т. 1. Общие положения: Учебник / Под общ. ред. д.ю.н., проф. Л.А. Новоселовой. - М.: Статут, 2017.

Приложение А

Образец данных из стартового датасета leaf_clust

NoEssai	NoBloc	Traitement	Annee	LatDD	LonDD	seFoliaire	Cultivar	Maturity5	RendVendable	RendPettit	RendMoy	RendGros	leaf_countNA	clr_N	clr_P	clr_K	clr_Ca	clr_Mg	clr_Fv
1	1	1	2003	46,75306	-72,3386	10% fleur	Goldrush	mid-season	18,9442	3,9458	11,4123	3,5861	0	0,332119	-2,51833	0,837938	-0,47947	-1,63908	3,466813
1	1	2	2003	46,75306	-72,3386	10% fleur	Goldrush	mid-season	40,3518	4,9159	17,5599	17,876	0	0,42523	-2,39496	1,138729	-0,51549	-2,34617	3,692658
1	1	3	2003	46,75306	-72,3386	10% fleur	Goldrush	mid-season	33,0379	6,322	13,6795	13,0364	0	0,445342	-2,36543	1,011227	-0,72908	-1,99074	3,62868
1	1	4	2003	46,75306	-72,3386	10% fleur	Goldrush	mid-season	37,5505	16,241	17,658	17,658	0	0,399326	-2,59918	0,794583	-0,62202	-1,57753	3,604817
1	1	5	2003	46,75306	-72,3386	10% fleur	Goldrush	mid-season	46,0089	4,6652	18,9987	22,245	0	0,651939	-2,46822	0,824243	-0,74708	-1,77507	3,514191
1	1	6	2003	46,75306	-72,3386	10% fleur	Goldrush	mid-season	41,0167	4,7742	17,1566	19,0859	0	0,488565	-2,52533	0,941888	-0,6426	-1,76319	3,900681
1	1	7	2003	46,75306	-72,3386	10% fleur	Goldrush	mid-season	46,7937	3,9785	22,2142	20,601	0	0,720917	-2,30125	0,720917	-0,87128	-1,70163	3,432323
1	1	8	2003	46,75306	-72,3386	10% fleur	Goldrush	mid-season	47,0117	5,0576	13,0582	28,9959	0	0,63795	-2,32136	0,817475	-0,80877	-1,81263	3,492956
1	2	1	2003	46,75306	-72,3386	10% fleur	Goldrush	mid-season	15,1183	4,1856	8,6001	2,326	0	0,472035	-2,28963	0,718928	-0,54183	-1,69826	3,337752
1	2	2	2003	46,75306	-72,3386	10% fleur	Goldrush	mid-season	36,2643	6,2457	19,2167	10,8019	0	0,34313	-2,2737	1,207536	-0,64145	-2,1645	3,528995
1	2	3	2003	46,75306	-72,3386	10% fleur	Goldrush	mid-season	44,8208	4,5017	15,9249	24,3942	0	0,542986	-2,28664	0,966427	-0,68594	-2,20326	3,664422
1	2	4	2003	46,75306	-72,3386	10% fleur	Goldrush	mid-season	43,7962	5,1557	14,4207	24,2198	0	0,555239	-2,43356	0,895232	-0,61055	-2,01235	3,605998
1	2	5	2003	46,75306	-72,3386	10% fleur	Goldrush	mid-season	42,9569	6,2784	16,1538	20,5247	0	0,569252	-2,42343	0,72147	-0,66109	-1,73028	3,524081
1	2	6	2003	46,75306	-72,3386	10% fleur	Goldrush	mid-season	49,2462	7,8698	24,0236	17,3528	0	0,725191	-2,32768	0,759722	-0,69452	-1,99121	3,528494
1	2	7	2003	46,75306	-72,3386	10% fleur	Goldrush	mid-season	41,8669	5,0467	16,4263	20,3939	0	0,507674	-2,50883	0,676348	-0,58817	-1,51729	3,430364
1	2	8	2003	46,75306	-72,3386	10% fleur	Goldrush	mid-season	49,5841	4,0875	22,127	23,3696	0	0,694266	-2,4233	0,781566	-0,74661	-1,82112	3,31552
1	3	1	2003	46,75306	-72,3386	10% fleur	Goldrush	mid-season	26,5524	3,9676	17,1113	5,4718	0	0,648465	-2,17599	1,11601	-0,69125	-2,40779	3,510555
1	3	2	2003	46,75306	-72,3386	10% fleur	Goldrush	mid-season	37,3761	5,8206	14,2681	17,2874	0	0,40998	-2,6619	1,025512	-0,4979	-1,85752	3,581825
1	3	3	2003	46,75306	-72,3386	10% fleur	Goldrush	mid-season	43,3275	4,8287	16,7642	21,7346	0	0,690962	-2,35356	1,091122	-0,65315	-2,45364	3,678269
1	3	4	2003	46,75306	-72,3386	10% fleur	Goldrush	mid-season	40,4935	6,0386	16,5462	17,9087	0	0,501083	-2,49215	0,702809	-0,57924	-1,63195	3,499443
1	3	5	2003	46,75306	-72,3386	10% fleur	Goldrush	mid-season	43,8616	6,976	19,075	17,8106	0	0,478408	-2,53362	0,801981	-0,58093	-1,66612	3,500278
1	3	6	2003	46,75306	-72,3386	10% fleur	Goldrush	mid-season	45,998	6,104	19,4674	20,4266	0	0,88388	-2,37221	0,658885	-0,71589	-1,92753	3,472871
1	3	7	2003	46,75306	-72,3386	10% fleur	Goldrush	mid-season	42,074	7,1504	22,3123	12,6113	0	0,54421	-2,47452	0,679705	-0,81422	-1,40532	3,470141
1	3	8	2003	46,75306	-72,3386	10% fleur	Goldrush	mid-season	54,5109	4,9922	24,4269	25,0918	0	0,595031	-2,35524	0,660659	-0,40405	-1,91341	3,417009
2	1	1	2003	46,75306	-72,3386	10% fleur	Eramosa	early	13,0255	4,7633	5,1775	3,0847	0	0,277111	-2,14225	0,070724	-0,23094	-1,92867	3,960033
2	1	2	2003	46,75306	-72,3386	10% fleur	Eramosa	early	28,4817	6,1258	10,3441	12,0118	0	0,364837	-2,44155	1,035486	-0,52196	-1,97724	3,540427
2	1	3	2003	46,75306	-72,3386	10% fleur	Eramosa	early	42,7825	4,578	14,8349	23,3696	0	0,442906	-2,53172	0,9807	-0,34766	-2,27389	3,729668
2	1	4	2003	46,75306	-72,3386	10% fleur	Eramosa	early	37,2671	4,9377	8,4039	23,9255	0	-0,00698	-2,30913	1,13601	-0,7351	-1,73106	3,646263
2	1	5	2003	46,75306	-72,3386	10% fleur	Eramosa	early	51,7423	2,8885	12,6004	36,2534	0	0,896348	-2,25617	1,321773	-1,42327	-2,36154	3,822854
2	1	6	2003	46,75306	-72,3386	10% fleur	Eramosa	early	34,5421	3,0629	11,5867	19,8925	0	-0,00824	-2,82005	0,895147	-0,4592	-1,17139	3,563728
2	1	7	2003	46,75306	-72,3386	10% fleur	Eramosa	early	50,9248	2,6596	9,156	39,1092	0	0,328549	-2,71716	1,213436	-0,34226	-2,05576	3,572201
2	1	8	2003	46,75306	-72,3386	10% fleur	Eramosa	early	41,747	6,3656	10,2569	25,1245	0	0,38908	-2,70361	1,028086	-0,3463	-1,94984	3,602593
2	2	1	2003	46,75306	-72,3386	10% fleur	Eramosa	early	12,2734	4,7306	4,905	2,6378	0	0,435647	-2,36062	1,114192	-0,46977	-2,24283	3,523378
2	2	2	2003	46,75306	-72,3386	10% fleur	Eramosa	early	40,3518	3,5088	10,4858	26,3562	0	0,094069	-2,52454	1,093855	-0,29095	-2,04762	3,675189
2	2	3	2003	46,75306	-72,3386	10% fleur	Eramosa	early	43,5455	5,2102	11,5322	26,8031	0	0,229728	-2,52427	1,089491	-0,33881	-2,06751	3,611366
2	2	4	2003	46,75306	-72,3386	10% fleur	Eramosa	early	53,2356	4,2292	13,625	35,3814	0	0,507605	-2,57934	1,089339	-0,50617	-1,99155	3,480112
2	2	5	2003	46,75306	-72,3386	10% fleur	Eramosa	early	46,4885	2,616	7,3684	36,5041	0	0,445739	-2,55878	1,329434	-0,49415	-2,39626	3,674016

Приложение Б

Пример входных данных для формирования предсказания на основе обученной модели.

X	clr_N	clr_P	clr_K	clr_Ca	clr_Mg	clr_Fv
7947	-0,95404	-0,69016	0,724968	1,9787	-0,2774	-2,02644
2523	1,7899	2,405679	-1,73385	-1,0709	-0,50633	-0,18933
1636	0,469605	-0,17158	-0,07758	-1,1131	0,260813	1,349168
8037	0,41677	1,382351	0,390128	-1,60608	-0,79764	0,718196
1273	-1,11509	-0,40832	-0,6423	0,263557	1,296465	0,362655
266	0,183471	-0,19645	0,717491	-0,96688	-0,68217	1,906689
7909	0,412028	0,123617	0,104434	-0,53009	0,047321	-0,16175
1882	0,133156	2,561225	-0,17449	-1,986	-0,86229	0,959315
2542	0,035363	-0,48032	1,150886	-0,03515	-0,69011	-0,03839
1838	0,939128	1,76771	-0,29156	-1,01733	-1,03989	0,172571
8089	-0,46231	-0,11121	-0,50103	0,346793	-0,00817	1,345678
871	0,368152	-0,39575	1,303962	-1,61155	-0,92409	2,499006
2299	-0,64941	0,233065	0,451427	1,13139	-0,65143	-1,18813
3024	0,784157	-0,19191	-0,19393	0,131969	0,072062	-0,73103
1186	-0,27561	2,223871	-0,80952	-1,10503	0,55388	-1,27861
1474	-0,57568	-0,34306	-0,29642	-0,38498	1,066299	0,450992
2527	-0,69477	-0,04466	-0,83549	0,579244	1,238239	-0,97752
8006	-0,08882	0,319771	0,775717	-0,39308	-0,97331	0,742236
1927	-1,10472	-2,29023	0,095086	1,901162	1,431488	-1,05218
1378	-1,35825	0,28598	-0,62298	-0,25541	1,828946	-0,90093
2931	0,205264	-0,523	-0,06412	0,485882	0,24244	-0,60908
1473	-1,015	-1,37486	-0,27307	0,429343	1,621128	0,221013
2398	0,927858	-1,84638	0,84092	-0,07018	0,163367	0,147023
1487	0,185648	0,286413	-0,98494	-1,23922	1,661982	-0,06125
1089	-0,03265	0,433286	0,311258	0,324645	-1,33021	0,921911
971	0,361742	0,635723	0,473139	-0,22417	-1,74386	1,586548
2412	0,185731	-0,66058	-0,04012	0,383107	0,568865	-0,91411
1343	-0,81092	0,766526	-0,53211	-0,20621	0,957215	-0,92789
3161	0,710713	-1,92119	0,348898	0,013529	-0,0192	1,946228
813	0,094362	1,329837	0,200738	-1,04784	0,050573	-1,28138
975	-0,96497	0,35383	0,495279	1,271538	-2,20145	2,181148
22	-1,1208	-1,90813	0,846952	0,824331	1,11231	-0,67062
3014	1,009934	-0,51273	-1,26773	0,789324	0,634851	-0,57582
2072	1,566074	0,893876	0,118476	-1,01743	-0,87967	-0,30304
2039	0,646943	-0,35287	-1,17097	0,748641	0,856619	-1,00738

	Maturity5_early.mid .season	Maturity 5_late	Maturity 5_mid.se ason	Maturity 5_mid.se ason.late
4	1	0	0	0
3	0	0	1	0
3	0	0	1	0
5	0	1	0	0
5	0	0	1	0
9	0	0	1	0
5	0	0	1	0
5	0	0	1	0
9	0	0	1	0
L	0	0	1	0
3	0	0	1	0
5	0	0	0	0
3	0	1	0	0
3	0	0	0	1
L	0	0	1	0
2	0	0	0	1
2	1	0	0	0
5	0	0	1	0
3	0	0	1	0
3	0	0	0	1
3	0	1	0	0
3	0	0	0	1
3	0	0	0	1
5	0	0	0	1
L	1	0	0	0
3	0	1	0	0
L	0	0	0	1
9	0	0	0	1
3	0	0	1	0
3	0	1	0	0
3	0	1	0	0
2	0	0	1	0
2	0	0	0	1
4	0	0	1	0
3	0	0	1	0
-	-	-	-	-

	Использовать фундаментальные результаты математики при создании моделей в области профессиональных интересов
Профессиональные компетенции	
ПК(У)-1	Уметь понять, интерпретировать поставленную задачу
ПК(У)-2	Уметь на основе анализа увидеть и корректно сформулировать результат
ПК(У)-3	Уметь грамотно пользоваться языком предметной области, как математическим, так и программным
ПК(У)-4	Способность решать задачи производственной и технологической деятельности на профессиональном уровне, включая: разработку алгоритмических и программных решений в области системного и прикладного программирования