

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«Тольяттинский государственный университет»

Институт математики, физики и информационных технологий
(наименование института полностью)

Кафедра «Прикладная математика и информатика»
(наименование кафедры)

УТВЕРЖДАЮ
Завкафедрой «Прикладная
математика и информатика»

_____ А.В. Очеповский
(подпись) (И.О. Фамилия)
« ____ » _____ 20__ г.

ЗАДАНИЕ
на выполнение бакалаврской работы

Студент Ламанов Вадим Вячеславович

1. Тема Построение классификаторов качества деталей на основе анализа экспериментальных данных ЗАО "Мегапласт"

2. Срок сдачи студентом законченной выпускной квалификационной работы
21 июня 2019 г.

3. Исходные данные к выпускной квалификационной работе:
экспериментальные данные по изготовлению пластиковых деталей литьем,
алгоритмы машинного обучения ID3, C4.5, CART.

4. Содержание выпускной квалификационной работы (перечень подлежащих
разработке вопросов, разделов)

Введение

1 Анализ состояния вопроса

2 Синтез классификаторов на основе алгоритмов индуктивного машинного обучения

3 Практическая реализация предложенных решений

Заключение

Список используемой литературы

5. Ориентировочный перечень графического и иллюстративного материала: набор формул, объясняющих математический аппарат алгоритма; результаты тестирования работы алгоритма на реальном наборе данных; диаграммы, поясняющие результат работы алгоритма; формы окна программы.

6. Консультанты по разделам _____

7. Дата выдачи задания «15» декабря 2018 г.

Заказчик, генеральный директор
ЗАО«Мегапласт»

Л.П. Платошин

Руководитель выпускной
квалификационной работы

В.С. Климов

Задание принял к исполнению

В.В. Ламанов

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

федеральное государственное бюджетное образовательное учреждение
высшего образования
«Тольяттинский государственный университет»

Институт математики, физики и информационных технологий

(наименование института полностью)

Кафедра «Прикладная математика и информатика»

(наименование кафедры)

УТВЕРЖДАЮ
Завкафедрой «Прикладная
математика и информатика»

_____ А.В. Очеповский
(подпись) (И.О. Фамилия)
«___» _____ 20__ г.

**КАЛЕНДАРНЫЙ ПЛАН
выполнения бакалаврской работы**

Студента Ламанова Вадима Вячеславовича

по теме «Построение классификаторов качества деталей на основе анализа экспериментальных данных ЗАО "Мегапласт"»

Наименование раздела работы	Плановый срок выполнения раздела	Фактически й срок выполнения раздела	Отметка о выполнении	Подпись руководителя
Анализ предметной области	1.10.2018	1.10.2018		
Синтез классификаторов на основе алгоритмов индуктивного машинного обучения	12.12.2018	12.12.2018		
Практическая реализация	17.01.2019	17.01.2019		

предложенных решений				
Подготовка пояснительной записки	25.02.2019	25.02.2019		
Подготовка презентации к защите	25.02.2019	25.02.2019		
Предзащита дипломной работы	10.06.2019	10.06.2019		
Проверка ВКР в системе «Антиплагиат. ВУЗ»	12.06.2019	12.06.2019		
Сдача пояснительной записки и отзыва на кафедру	17.06.2019	17.06.2019		
Защита дипломной работы	1.07.2019	1.07.2019		

Руководитель выпускной
квалификационной работы

В.С. Климов

Задание принял к исполнению

В.В. Ламанов

АННОТАЦИЯ

Тема бакалаврской работы: «Построение классификаторов качества деталей на основе анализа экспериментальных данных ЗАО "Мегапласт"».

В данной бакалаврской работе исследуется возможность применения алгоритмов построения деревьев принятия решения для создания классификаторов качества пластиковых деталей, выполняемых литьем под давлением.

В работе предложен алгоритм построения классификатора качества пластиковых деталей с использованием алгоритмов индуктивного машинного обучения – ID3, C4.5, CART. Полученные классификаторы можно применять для сквозной диагностики качества изготавливаемых пластиковых деталей.

При подготовке исследований, представленных в рамках данной бакалаврской работы, была опубликована 1 статья.

Бакалаврская работа состоит из введения, трёх глав, заключения и списка литературы.

Во введении описывается актуальность проводимого исследования, дается краткая характеристика проделанной работы.

В первой главе проводится обзор предметной области представленного исследования. Описываются исходные данные для построения классификаторов качества.

Во второй главе описывается математический аппарат алгоритмов индуктивного машинного обучения ID3, C4.5, CART. Формулируется алгоритм построения классификаторов с использованием имеющихся данных.

В третьей главе описывается программная реализация предложенных подходов.

В заключении представлены выводы по проделанной работе.

В работе присутствуют 2 таблицы, 11 рисунков. Список литературы состоит из 20 литературных источников. Общий объем выпускной квалификационной работы составляет 40 страниц.

ABSTRACT

The subject of bachelor's work: "The application of the method of decision trees for data classification."

In this bachelor's work, the possibility of using decision-making algorithms to create quality classifiers for plastic parts performed by injection molding is being investigated.

The paper proposes an algorithm for constructing a quality classifier for plastic parts using inductive machine learning algorithms - ID3, C4.5, CART. The resulting classifiers can be used for end-to-end diagnostics of the quality of manufactured plastic parts.

In preparing the studies presented in this bachelor's work, 1 article was published.

Bachelor's work consists of introduction, three chapters, conclusion and list of references.

The introduction describes the relevance of the study, provides a brief description of the work done.

The first chapter provides an overview of the subject area of the presented research. The source data for the construction of quality classifiers are described.

The second chapter describes the mathematical apparatus of the algorithms of inductive machine learning ID3, C4.5, CART. An algorithm for constructing classifiers using the available data is formulated.

The third chapter describes the software implementation of the proposed approaches.

In the conclusion presents conclusions on the work done. In the work there are 88 tables, 88 figures. References consists of 20 references. The total amount of final qualifying work is 40 pages.

ОГЛАВЛЕНИЕ

1 ПРИМЕНЕНИЕ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПОСТРОЕНИЯ КЛАССИФИКАТОРОВ.....	7
2 РАЗРАБОТКА КЛАССИФИКАТОРА КАЧЕСТВА ДЕТАЛЕЙ НА ОСНОВЕ ДЕРЕВЬЕВ ПРИНЯТИЯ РЕШЕНИЙ.....	12
2.1 Постановка задача классификации	12
2.2 Алгоритмы построения деревьев принятия решений	13
2.2 Анализ исходных данных для построения классификатора качества деталей.....	22
2.3 Обзор полученных классификаторов и сравнение их характеристик.....	25
3 ПРАКТИЧЕСКАЯ РЕАЛИЗАЦИЯ ПРЕДЛОЖЕННЫХ РЕШЕНИЙ.....	32
3.1 Описание разработанного программного обеспечения	32
3.2 Пример использования классификаторов	34
ЗАКЛЮЧЕНИЕ	36
СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ	37

ВВЕДЕНИЕ

Решение проблем диагностики и управление технологическими процессами, которые возникают на различных производствах, часто рассматривается как задача классификация. Например, имеется набор косвенных измеренных, в процессе изготовления детали, параметров по которым необходимо оценить класс ее качества. Использование стандартных подходов математического и физического моделирование заключается в объяснении зависимости между классом качества детали и измеряемыми параметрами. Такой подход на практике сложнореализуем, так как требует привлечение различных специалистов как из области теории автоматического управления, так и из области физического моделирования.

Развитие технологий машинного обучения позволило упростить решение таких задач, за счет разработки алгоритмов автоматизированного синтеза классификаторов, основанного на анализе экспериментальных данных технологического процесса.

В качестве моделей классификации могут применяться такие структуры, как деревья принятия решений. При этом их построение возможно с использованием таких алгоритмов как ID3, C4.5, CART.

Исследованиями в области практического применения данных алгоритмов занимаются такие ученые, как Fachao Li, Dandan Jiang, Lin Zhu, Yang Yang, Hiroshi Narazaki, Ichiro Shigaki, Jeanette Auer, Richard Hall, Fan Min, William Zhu, Jyh-Yeong Chang, Chien-Wen Cho, Su-Hwang Hsieh, Shi-Tsung Chen, Łukasz Bartczuk, Danuta Rutkowska, Didi Jia, Wu Xie, Zhiyong Chen, Baohua Qiang и др [1-20].

В данной бакалаврской работе рассматривается проблема диагностики качества пластиковых деталей выполненных литьем под давлением. На качество изготовления пластиковых деталей оказывает влияние множество параметров. Среди них - температура массы, температура сушки, температура зоны загрузки, давление и температура сушки. Существуют

рекомендации с указанием диапазонов значений этих параметров в зависимости от типа используемого пластика. Существующие проблемы заключаются в следующем:

- в рекомендуемых параметрах не учитывается геометрическая форма изготавливаемых деталей и за-за этого возникают дефекты в деталях;
- отклонение одних температурных параметров может быть компенсировано отклонениями других параметров, поэтому даже при выходе значений из рекомендованного диапазона в деталь будет качественной.

Поэтому актуальной задачей является разработка алгоритмов для автоматизированного контроля качества деталей.

Целью работы является – построение классификаторов для диагностики качества литья пластиковых деталей с использованием алгоритмов машинного обучения (ID3, C4.5, CART).

В бакалаврской работе на основе экспериментальных данных, предоставленных предприятием ЗАО Мегапласт, с использованием алгоритмов машинного обучения были построены классификаторы для определения класса качества литья. Также полученные классификаторы (деревья решений) тестировались для оценки точности их работы.

В ходе выполнения бакалаврской работы была спроектировано и разработано программное обеспечения реализующее процесс построения деревьев принятия решений.

1 ПРИМЕНЕНИЕ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПОСТРОЕНИЯ КЛАССИФИКАТОРОВ

Создание алгоритма диагностики для диагностики кого-либо технологического процесса связано с решением задачи классификации, когда на основе измеренных параметров требуется определить в общем случае класс возникающей ситуации, в частном случае – класс качества изготовленной детали. В самом простом случае при диагностике технологического процесса решается задача бинарной классификации, когда классов всего два – «качественная деталь», «брак».

Алгоритм, который позволяет на основе измеренных параметров определять класс качества детали принято называть классификатором. В классическом понимании, для построения классификатора качества какого-либо технологического процесса требуется привлечение значительного количества экспертов, в том числе и технолога, задача которого объяснить, как влияют параметры на получаемое качество изделий.

В настоящее время, задачу построения классификатора можно решать другим путем - за счет привлечения алгоритмов машинного обучения. Алгоритмы машинного обучения позволяют автоматизировано генерировать классификаторы, на основе экспериментальных данных технологического процесса. Причем современные алгоритмы машинного обучения самостоятельно определяют значимость каждого параметра, и осуществлять поиск закономерностей между параметрами и классом качества получаемой детали.

Один из вариантов представления классификатора – в виде дерева принятия решений. К современным алгоритмам построения деревьев принятия решений относятся ID3, C4.5 и CART.

Эти алгоритмы имеют разный математический аппарат и получаемые с помощью них классификаторы с использованием одних и тех же данных тоже существенно различаются. Примечательно, что заранее никогда не

известно с помощью кого из этих алгоритмов удастся получить классификатор с наиболее удачным набором характеристик. Поэтому необходимо экспериментальным способом определять наиболее подходящий для данного набора данных алгоритм.

Одна из проблем производства пластиковых деталей методом литья под давлением является сложность соблюдения требуемых температурных режимов. При литье деталей существует большое количество внешних (неконтролируемых) параметров, влияющих на отклонение температурных режимов. Например, температура окружающей среды которая может значительно меняться в течении суток, меняющийся темп производства: высокий темп производства не дает остыть оснастке, что приводит к перегреву деталей, низкий темп производства вынуждает прогревать оснастку перед началом литья, при этом темп производства может меняться в течение дня.

Хоть производители пластиковых материалов и разработали таблицу с рекомендуемыми параметрами для литья пластмасс под давлением, но рекомендуемые параметры не учитывает геометрическую форму деталей. Это означает, что при соблюдении данных параметров в тонких местах детали при ее изготовлении возможно появление трещин или других термических повреждений из-за избыточной температуры, а в толстых местах детали возможно появление недоливов из-за недостаточного прогрева массы во время литья (рисунок 1.1). Уточнение температурных режимов производстве производится опытным путем. Т.е. литье повторяют при различных температурных режимах, пока не будет достигнута приемлемая стабильность качества литья.

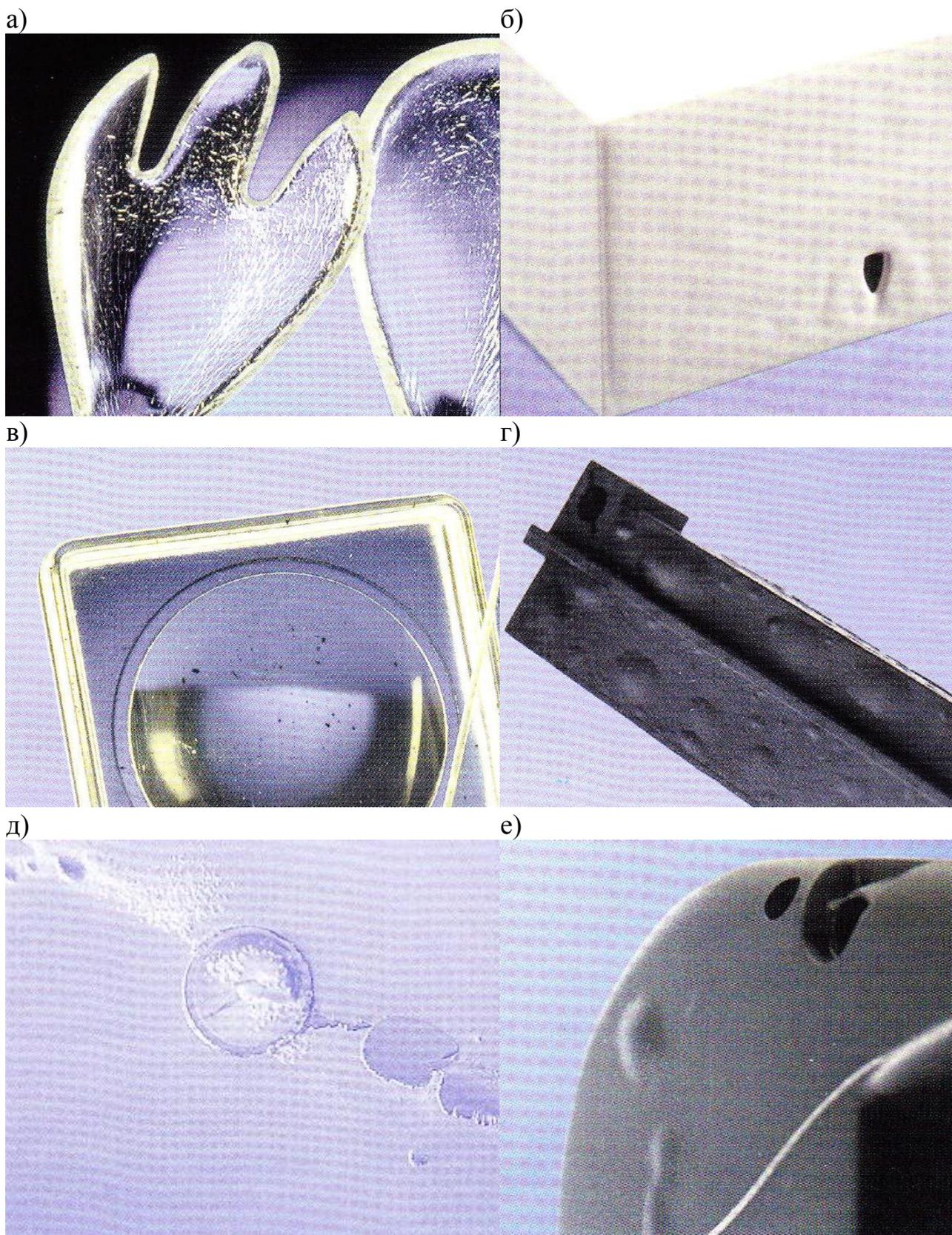


Рисунок 1.1 – Дефекты при литье деталей: а – трещины; б – недолив; в – термические повреждения г – разложение поверхности; д – внутренние пузыри; е – вздутие.

Также необходимо учитывать, что отклонение одних температурных параметров может быть компенсировано отклонениями других параметров, поэтому даже при выходе значений из рекомендованного диапазона деталь на выходе может быть качественной.

В процессе уточнения параметров подбора температурных режимов для литья на производстве скапливается большое количество экспериментальных данных. Данные одного эксперимента объединяют в себе набор температурных характеристик и данные о полученном качестве детали.

В данной бакалаврской работе использован набор таких экспериментальных данных, предоставленный предприятием ЗАО «Мегапласт» для облицовки левой стайки автомобиля Largus, выпаленной из АБС пластика.

В настоящий момент на ЗАО «Мегапласт» контроль качества деталей осуществляется путем их визуального осмотра. Визуальный контроль обладает следующими недостатками:

- высокая трудоемкость;
- эффективность визуального контроля зависит от квалификации;
- низкая производительность контроля.

Поэтому актуальной проблемой является автоматизация процесса диагностики деталей, выполняемых литьем под давлением.

Предполагается, что данную проблему можно решить, найдя зависимость между измеряемыми в процессе литья температурными характеристиками и классом качества получаемых деталей. Для получения данных зависимостей предлагается использовать алгоритмы машинного обучения ID3, C4.5 и CART. В качестве исходных данных будут использованы экспериментальные данные по литью левой стайки автомобиля Largus из АБС пластика, предоставленных ЗАО «Мегапласт»

Таким образом, актуальной можно признать цель данного исследования – синтез классификатора качества пластиковых изделий на основе анализа экспериментальных данных.

Поставленную цель предполагается достигнуть путем решения следующих задач:

1. Анализ технологии диагностики качества пластиковых деталей, выполняемых с помощью литья под давлением.

2. Представление процесса диагностики как задачи классификации данных и синтез классификаторов качества на основе алгоритмов машинного обучения (ID3, CART, C4.5).

3. Тестирование классификаторов качества и определения их точности работы.

4. Разработка программного обеспечения для автоматизации процесса построения классификаторов.

2 РАЗРАБОТКА КЛАССИФИКАТОРА КАЧЕСТВА ДЕТАЛЕЙ НА ОСНОВЕ ДЕРЕВЬЕВ ПРИНЯТИЯ РЕШЕНИЙ

2.1 Постановка задачи классификации

Формально задача классификации описывается следующим образом. Существует множество входных переменных X – описаний объектов, а также множество Y (конечное) обозначений классов. В задаче классификации существует неизвестная целевая зависимость – отображение $y^*: X \rightarrow Y$, значения которой известны только на объектах из обучающей выборки X^m :

$$X^m = (x_1, y_1), \dots, (x_m, y_m) , \quad (2.1)$$

где m – размер обучающей выборки, x_i – вектор значений входных переменных для i -го объекта из обучающей выборки, y_i – значение метки класса для i -го объекта из обучающей выборки. При этом вектор x_i описывается набором значений атрибутов A , а y_i может принимать одно значений меток класса C_1, C_2, \dots, C_k

$$\begin{cases} x_i = (A_1, A_2, \dots, A_n) \\ y_i \in C_1, C_2, \dots, C_k \end{cases} , \quad (2.2)$$

где, n – количество атрибутов, описывающих объект из обучающей выборки, k – количество меток класса в обучающей выборке.

Требуется построить алгоритм $\alpha: X \rightarrow Y$, способный классифицировать объект описываемый вектором входных значений x , при $x \in X$.

Синтез алгоритма диагностики качества пластиковых деталей, выполняемых литьем под давлением можно рассматривать как задачу построения классификатора, где $y_i \in C_1, C_2, \dots, C_k$ - класс качества детали, а $x_i = (A_1, A_2, \dots, A_n)$ - измеряемые в процессе литья детали параметры.

Одним из возможных представлений модели классификации данных является дерево принятия решений. Дерево решений – это связный ациклический граф, представляющий правила классификации в иерархической последовательной структуре. Дерево состоит узлов и листов.

В каждом узле расположено проверка по одному из атрибутов вектора x_i . В каждом листе расположено одно из значений меток класса C_1, C_2, \dots, C_k . Чтобы классифицировать любой из объектов обучающей выборки надо спуститься по дереву до листа и выдать значение метки класса листа. В зависимости от значений атрибутов классифицируемого объекта спускаясь по дереву можно прийти к различным листам.

2.2 Алгоритмы построения деревьев принятия решений

Общий принцип построения деревьев решений, заключается в рекурсивном разбиении множества объектов T из обучающей выборки на подмножества, содержащие объекты, относящиеся к одинаковым классам. Блок-схема данного алгоритма показана на рисунке 2.1. Относительно обучающей выборки T и множества классов C_1, C_2, \dots, C_k возможны три ситуации:

- множество T содержит один или более объектов, относящихся к одному классу C_i . Тогда дерево решений для T - это лист, определяющий класс C_i ;
- множество T не содержит ни одного объекта (пустое множество). Тогда данное множество T – это лист, и класс, ассоциированный с листом, выбирается из другого множества, отличного от T , например из множества, ассоциированного с родителем;
- множество T содержит объекты, относящиеся к разным классам. В этом случае следует разбить множество T на некоторые подмножества. Для этого выбирается один из атрибутов A_1, A_2, \dots, A_k и T разбивается на подмножества T_1, T_2, \dots, T_j , по выбранному атрибуту и соответствующему ему условию.

Эти шаги повторяются рекурсивно для всех получающихся подмножеств до тех пор, пока все подмножества не будут объявлены листьями.



Рисунок 2.1 –Алгоритм построения дерева решений

Все алгоритмы автоматического построения деревьев принятия решений относятся к индуктивным алгоритмам машинного обучения (и к подобласти “обучения с учителем”).

Одним из важных вопросов построения дерева принятия решений является выбор таких условий в узлах, которые обеспечат наилучшее разбиение обучающей выборки на подмножества. Это необходимо, чтобы получившиеся дерево было, во-первых компактным (содержало наименьшее количество узлов), а во-вторых обеспечивало максимальную точность классификации.

Алгоритм ID3 для построения дерева принятия решения выбирает из всех возможных вариантов такую проверку в узле, которая обеспечивают максимальное снижение информационной двоичной энтропии и с учетом равномерного распределения объектов по подмножествам

Рассмотрим этот вопрос подробнее на множестве T объектов, содержащих k меток класса – C_1, C_2, \dots, C_k . Если перемешать объекты из множества T и расположить их в ряд, то мы реализуем операцию под названием "перестановка". При этом считается, что объекты с одинаковыми метками класса идентичны. Тогда количество уникальных перестановок для множества рассчитывается, как:

$$W = \frac{N!}{N_1! \cdot N_2! \cdot \dots} = \frac{N!}{\prod_{i=1}^k N_i!}, \quad (2.3)$$

где N_i – количество объектов с меткой класса C_i в множестве T , N – количество объектов в обучающей выборке.

Если все уникальные перестановки пронумеровать числами от 0 до $W-1$, то количество бит необходимое для кодирования каждого уникального варианта перестановки равно $\log_2(W)$. Среднее количество бит на каждую перестановку называется комбинаторной энтропией и рассчитывается как:

$$S = \frac{\log_2(W)}{N} = \frac{1}{N} \cdot \log_2\left(\frac{N!}{\prod_{i=1}^k N_i!}\right) \quad (2.4)$$

Важное комбинаторной энтропии, которое используется при построении дерева принятия решения – чем меньше энтропия, тем однороднее множество. Самые однородные множества в дереве принятия решений – это листы не содержащие примесей других классов. Самое неоднородное множество – начальная обучающая выборка.

Для удобства расчета энтропии выражение, приведенное выше, можно упростить, воспользовавшись формулой Стирлинга:

$$\ln N! = N \cdot \ln N - N + O(\ln N) \approx N \cdot \ln N - N \quad (2.5)$$

Тогда энтропию для множества T с учетом упрощений можно рассчитать так:

$$S = -\sum_{i=1}^k \left(\frac{N_i}{N} \cdot \log_2 \frac{N_i}{N}\right) \quad (2.6)$$

С учетом того, что значение энтропии S зависит от рассматриваемого множества T , выразим ее как функцию:

$$S(T) = -\sum_{i=1}^k \left(\frac{\text{freq}(T, C_i)}{|T|} \cdot \log_2 \frac{\text{freq}(T, C_i)}{|T|}\right), \quad (2.7)$$

где $\text{freq}(T, C_i)$ – функция, которая возвращает количество объектов с меткой класса C_i в множестве T .

При построении дерева принятия решения в каждый узел помещается условие, которое разбивает исходное множество T на несколько j подмножеств:

$$T = T_1 \cup T_2 \cup \dots \cup T_j \quad (2.8)$$

Тогда для каждого подмножества T_1, T_2, \dots, T_j по формуле 2.7 можно рассчитать энтропию.

Таким образом энтропию множества T после разбиения на подмножества рассчитывается как:

$$S_0(T, T_1, T_2, \dots, T_j) = \sum_{i=1}^j \left(\frac{|T_i|}{|T|} \cdot S(T_i) \right) \quad (2.9)$$

Если из энтропии $S(T)$ вычесть энтропию разбиения $S_0(T, T_1, T_2, \dots, T_j)$, то можно рассчитать прирост информации $G(T, T_1, T_2, \dots, T_j)$, обеспечивающийся данным условием в узле:

$$G(T, T_1, T_2, \dots, T_j) = S(T) - \sum_{i=1}^j \left(\frac{|T_i|}{|T|} \cdot S(T_i) \right) \quad (2.10)$$

Чем однородней получаются подмножества в результате разбиения, тем больший прирост информации обеспечивает данное разбиение.

Таким образом, алгоритм ID3 на каждом шаге выполнения выбирает такую проверку в узле, которая обеспечивает максимальный прирост информации:

$$G(T, T_1, T_2, \dots, T_j) \rightarrow \max \quad (2.11)$$

Энтропия будет равна нулю, если в результате построения дерева принятия решения удалось достигнуть листов без содержания примесей посторонних классов.

Для каждого категориального атрибута существует только один вариант разбиения – когда исходное множество T делится на подмножества в количестве равном значений данного атрибута.

Для каждого числового атрибута существует несколько вариантов разбиений, количество которых равно мощности множества порогов. Множество порогов находится путем записи всех уникальных значений данного числового атрибута в обучающей выборке в порядке возрастания без повторов с отбрасыванием наибольшего значения. Затем каждое значение порога z выступает в роли точки деления исходного множества T , состоящего из объектов t , на подмножества T_1 и T_2 (2.12):

$$\begin{cases} T_1 = \{ t \in T \mid a(t) \leq z \\ T_2 = \{ t \in T \mid a(t) > z \} \end{cases} \quad (2.12)$$

где, a – значение атрибута, по которому производится разбиение.

В конечном итоге из всех возможных вариантов разбиения выбирается то, которое соответствует критерию (2.11). Как было сказано выше, построение дерева ведется до тех пор, пока не будут найдены все листы.

Основной недостаток алгоритма ID3 — тенденция к переобучению, которая заключается в том, что алгоритм склонен при рассмотрении вариантов разбиений к выбору атрибутов с наибольшим количеством уникальных значений. Действительно, если в результате разбиения будут получены множества, содержащие по одному объекту, образующему один класс, то частота $freq(C_j, T)$ появления этого класса окажется равной числу примеров, то есть 1 (и $|T|=1$). Таким образом, в выражении для оценки количества информации появится $\log_2(1) = 0$. Следовательно, и все выражение обратится в ноль, то есть $Info_s(T) = 0$. Тогда $Gain(S) = Info(T) - Info_s(T)$ будет выглядеть, как $Gain(S) = Info(T) - 0$. Всегда будет обеспечен максимальный прирост информации $Gain(S) = \max$, что приведет к выбору алгоритмом соответствующего атрибута.

Данная проблема в алгоритме C4.5 решается введением нормировки. В рассмотрение включается дополнительный показатель, который представляет собой оценку потенциальной информации, созданной при разбиении множества T на n подмножеств T_i :

$$SplitInfo(S) = - \sum_{i=1}^n (|T_i|/|T|) \cdot \log_2(|T_i|/|T|) \quad (2.13)$$

С помощью этого показателя можно модифицировать критерий прироста информации, перейдя к отношению:

$$GainRatio(S) = G(S) / SplitInfo(S) \quad (2.14)$$

Новый критерий позволяет оценить долю информации, полученной при разбиении, которая является полезной, то есть способствует улучшению

классификации. Использование данного отношения обычно приводит к выбору более удачного атрибута, чем обычный критерий прироста.

Помимо алгоритма ID3 и C4.5, для построения деревьев принятия решений существует также алгоритм CART.

Если сравнить алгоритмы ID3 и C4.5, с алгоритмом CART то отличия последнего будут следующие:

- бинарное представление дерева решений;
- другой параметр оценки качества разбиения;
- наличие механизма отсечения ветвей дерева;
- другой алгоритм обработки пропущенных значений;
- возможность выполнения регрессионного анализа.

Алгоритм CART строит бинарные деревья – это значит, что каждый всегда имеет двух потомков. Потомок для которого выполняется правило в узле называют правым (right), а в котором не выполняется – левым (left).

При выборе условия помещаемого в узел дерева интуитивно понятно, что получаемые в результате разбиения подмножества должны содержать как можно меньше примесей из других классов (т.е. быть как можно “чище”). Ведь если мы получаем максимально “чистое” множество, все записи в котором относятся к одному классу, то это множество можно объявить листом. В алгоритме CART эта идея формализована посредством индекса Gini:

$$Gini(T) = 1 - \sum_{i=1}^n p_i^2, \quad (2.15)$$

где T – множество после разбиения, содержащее n классов, p_i – относительная частота i -го класса в T .

Если в результате разбиения получается подмножество, не содержащее в себе примесей других классов, то $n=1$, $p_1=1$, следовательно, индекс $Gini(T)=0$.

Так как деревья CART – бинарные, то в результате разбиения S всегда получается два подмножества (T_1 и T_2) и показатель качества разбиения в этом случае будет зависеть от “чистоты” обоих множеств ($Gini(T_1)$ и $Gini(T_2)$)

с учетом доли их участия при разбиении ($\frac{|T_1|}{|T|}$ и $\frac{|T_2|}{|T|}$):

$$Gini_s(T) = \frac{|T_1|}{|T|} Gini(T_1) + \frac{|T_2|}{|T|} Gini(T_2) \quad (2.16)$$

Наилучшим считается разбиение, для которого $Gini_s(T)$ минимально, т.е. когда оба получившихся при разбиении подмножества максимально “чистые”.

$$Gini_s(T) \rightarrow \min \quad (2.17)$$

Обозначим через N число примеров в узле-предке, через L и R — число примеров в левом и правом потомке соответственно, l_i и r_i — число экземпляров i -го класса в левом/правом потомке.

Тогда качество разбиения оценивается по следующей формуле:

$$Gini_s = \frac{L}{N} \left(1 - \sum_{i=1}^n \left(\frac{l_i}{L}\right)^2\right) + \frac{R}{N} \left(1 - \sum_{i=1}^n \left(\frac{r_i}{R}\right)^2\right) \rightarrow \min \quad (2.18)$$

Упростим выражение:

$$Gini_s = \frac{1}{N} \left(L \left(1 - \frac{1}{L^2} \sum_{i=1}^n l_i^2\right) + R \left(1 - \frac{1}{R^2} \sum_{i=1}^n r_i^2\right) \right) \rightarrow \min \quad (2.19)$$

Фактически нам необходимо решить следующую задачу:

$$L \left(1 - \frac{1}{L^2} \sum_{i=1}^n l_i^2\right) + R \left(1 - \frac{1}{R^2} \sum_{i=1}^n r_i^2\right) \rightarrow \min \quad (2.20)$$

Если раскрыть скобки, то мы получим:

$$L - \frac{1}{L} \sum_{i=1}^n l_i^2 + R - \frac{1}{R} \sum_{i=1}^n r_i^2 \rightarrow \min \quad (2.21)$$

Зная, что $N=L+R$, получим:

$$N - \frac{1}{L} \sum_{i=1}^n l_i^2 - \frac{1}{R} \sum_{i=1}^n r_i^2 \rightarrow \min \quad (2.22)$$

Фактически решаемая задача сводится к:

$$-\frac{1}{L} \sum_{i=1}^n l_i^2 - \frac{1}{R} \sum_{i=1}^n r_i^2 \rightarrow \min \quad (2.23)$$

Или, если умножить обе части на -1:

$$\frac{1}{L} \sum_{i=1}^n l_i^2 + \frac{1}{R} \sum_{i=1}^n r_i^2 \rightarrow \max \quad (2.24)$$

На практике для удобства расчета качества разбиения в алгоритме CART используется обобщённый показатель $Q(s|t)$, где s — идентификатор разбиения, t — идентификатор узла:

$$Q(s|t) = 2P_L P_R \sum_{j=1}^N (P(j|t_L) - P(j|t_R)), \quad (2.25)$$

Где t_L и t_R — левый и правый потомки узла соответственно; $P_L = K_L / K$ — отношение числа примеров в левом потомке узла t к общему числу примеров; $P_R = K_R / K$ — отношение числа примеров в правом потомке узла t к общему числу примеров; $P(j|t_L) = K_j^L / K_t$ — отношение числа примеров j -го класса в t_L к общему числу примеров в t_L ; $P(j|t_R) = K_j^R / K_t$ — отношение числа примеров j -го класса в t_R к общему числу примеров в t_R .

Тогда наилучшим разбиением в узле t будет то, которое максимизирует показатель:

$$Q(s|t) \rightarrow \max \quad (2.26)$$

Результаты сравнения алгоритмов построения деревьев принятия решений представлены в таблице 1.

Таблица 1 – Сравнительная таблица алгоритмов построения деревьев принятия решений

Показатель	Алгоритм построения дерева решения		
	ID3	C4.5	CART
Получаемые модели	Модель классификации	Модель классификации	Модели классификации и регрессии
Возможное количество потомков у узла	≥ 2	≥ 2	всегда 2 (бинарное дерево)
Критерий для выбора используемого разбиения	Прирост информации G	Нормированный прирост информации $GainRatio$	Индекс Джини $Gini$
Учитываемые показатели при выборе условия разбиения для узла	<ul style="list-style-type: none"> • количество в потомках узла примесей других классов 	<ul style="list-style-type: none"> • количество в потомках узла примесей других классов • равномерность распределения объектов по подмножествам 	<ul style="list-style-type: none"> • количество в потомках узла примесей других классов

2.2 Анализ исходных данных для построения классификатора качества деталей

Предприятие ЗАО «Мегапласт» предоставило экспериментальные данные по литью детали из АБС пластика – облицовка левой стоки внутри салона для автомобиля Largus. Данные содержат результаты литья 99 деталей

при различных режимах. При сборе этих экспериментальных данных производилось измерение следующих параметров:

- A_1 – температура массы пластика, °С;
- A_2 – температура формы, °С;
- A_3 – температура зоны загрузки, °С;
- A_4 – давление, бар;
- A_5 – температура сушки, °С.

Выходным параметром являлся класс y качества деталей. В нашем случае использовано два класса качества: первый класс с меткой «С1» – деталь получена без дефектов литья, второй класс с меткой «С2» – деталь один или несколько дефектов литья. Схематично исходные данные показаны на рисунке 2.2. Фрагмент экспериментальных данных представлен в таблице 2.

$$X^m = \{(x_1, y_1), \dots, (x_m, y_m)\} \quad m=99 \quad i=1 \dots m$$



Рисунок 2.2 – Исходные данные для построения классификатора

Таблица 2 – Обучающая выборка

№	A ₁	A ₂	A ₃	A ₄	A ₅	C
1	214	64	36	107	71	C2
2	248	85	60	106	73	C2
3	247	56	54	134	88	C2
4	237	31	35	146	81	C2
5	258	44	47	85	89	C2
6	266	44	57	116	86	C2
7	238	35	49	80	80	C2
8	244	61	44	106	90	C1
9	253	43	41	117	77	C1
10	227	84	42	137	79	C2
11	263	79	35	108	86	C2
12	240	68	45	92	70	C2
13	279	55	45	133	88	C2
14	213	40	45	125	83	C2
15	231	48	56	60	74	C2
16	267	56	37	83	84	C1
17	259	90	53	121	86	C2
18	261	76	44	125	81	C2
18	279	46	43	117	75	C2
19	239	30	37	155	78	C2
20	273	64	59	116	86	C2
21	267	78	35	75	88	C2
22	204	47	50	100	80	C1
23	260	84	57	121	79	C2
24	275	71	47	106	89	C2
25	273	66	48	108	87	C2
26	254	80	56	62	78	C2
27	219	61	37	141	90	C2
28	233	36	49	110	72	C2
29	241	48	39	65	76	C1
30	215	65	46	101	86	C2
31	259	52	45	71	83	C1
32	252	56	54	72	85	C2
33	220	53	47	85	85	C2
34	243	54	52	147	76	C1
35	276	51	59	64	74	C2
36	266	54	51	99	87	C1
37	214	64	36	107	71	C2
...		
99	253	47	58	111	80	C1

2.3 Обзор полученных классификаторов и сравнение их характеристик

По данным представленным в таблице 2 были получены классификаторы качества литья в виде деревьев принятия решений. Данные деревья позволяют отделить случаи появления дефектов (класс «С2», на рисунках отмечен красным цветом) от случаев качественного литья (класс «С1», на рисунках отмечен синим цветом). При построении деревьев использовались различные алгоритмы индуктивного машинного обучения, и полученные результаты представлены ниже на рисунках:

- дерево решений, полученное с использованием алгоритма ID3, представлено на рисунке 2.3;
- дерево решений, полученное с использованием алгоритма C4.5, представлено на рисунке 2.4;
- дерево решений, полученное с использованием алгоритма CART, представлено на рисунке 2.5.

Для того чтобы определить, какой из полученных классификаторов является лучшим необходимо проанализировать их структуру и точностные характеристики.

При оценке структуры классификаторов необходимо учесть количество узлов и листов, чтобы понять насколько общими являются правила, входящие в состав дерева принятия решений.

Для оценки точности классификаторов необходимо проверить их работы с помощью различных методов. В нашем случае точность классификаторов будет оцениваться с использованием методов:

- методом перекрёстной проверки при различном количестве k частей (при k равном 3; 5 и 7);
- методом Leave-One-Out.

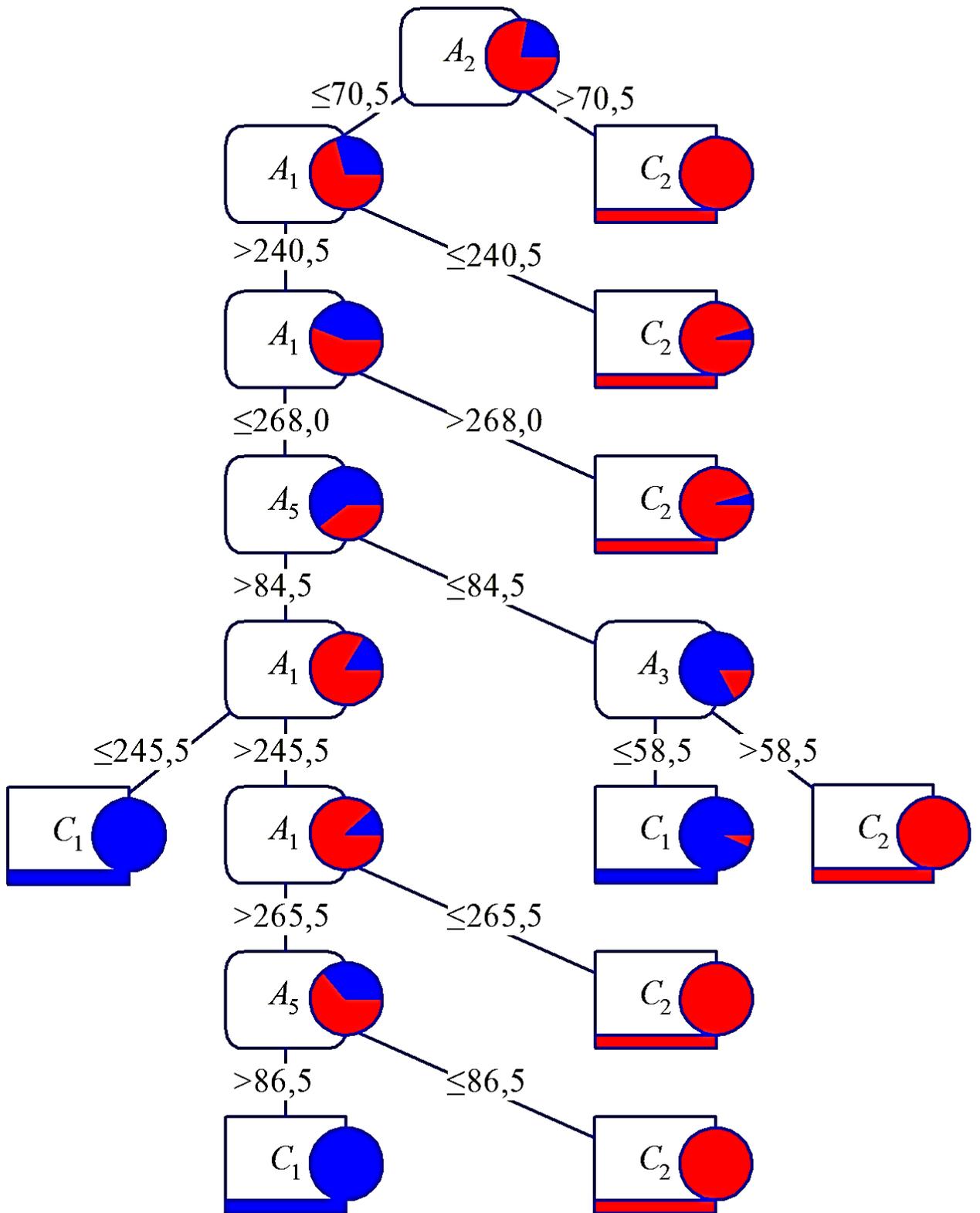


Рисунок 2.3 – Дерево классификации, полученный с помощью алгоритма ID3 (синим цветом отмечен класс C_1 «качественное литье», синим цветом - класс C_2 «брак»)

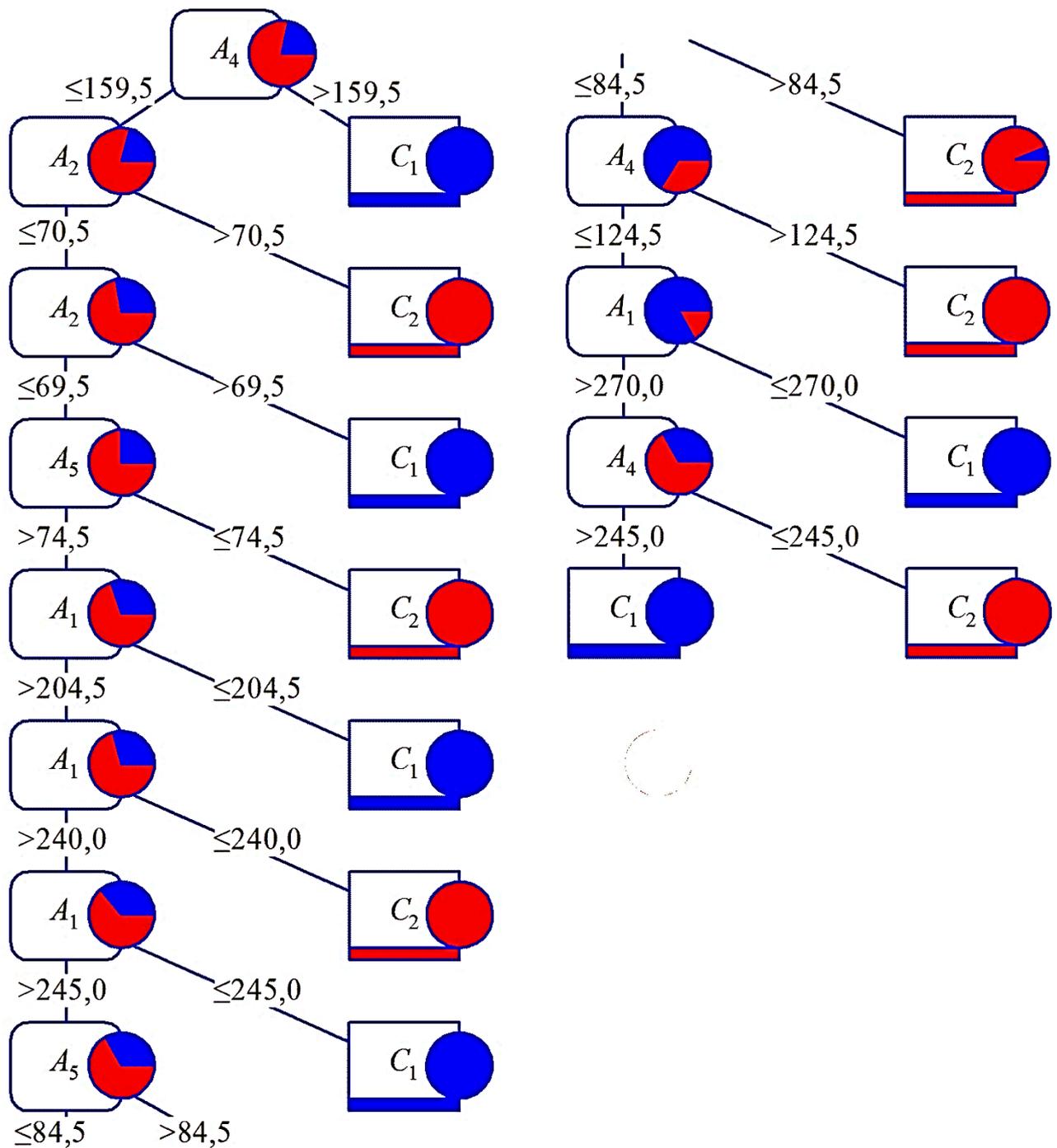


Рисунок 2.4 – Дерево классификации, полученный с помощью алгоритма C4.5 (синим цветом отмечен класс C1 «качественное литье», синим цветом - класс C2 «брак»)

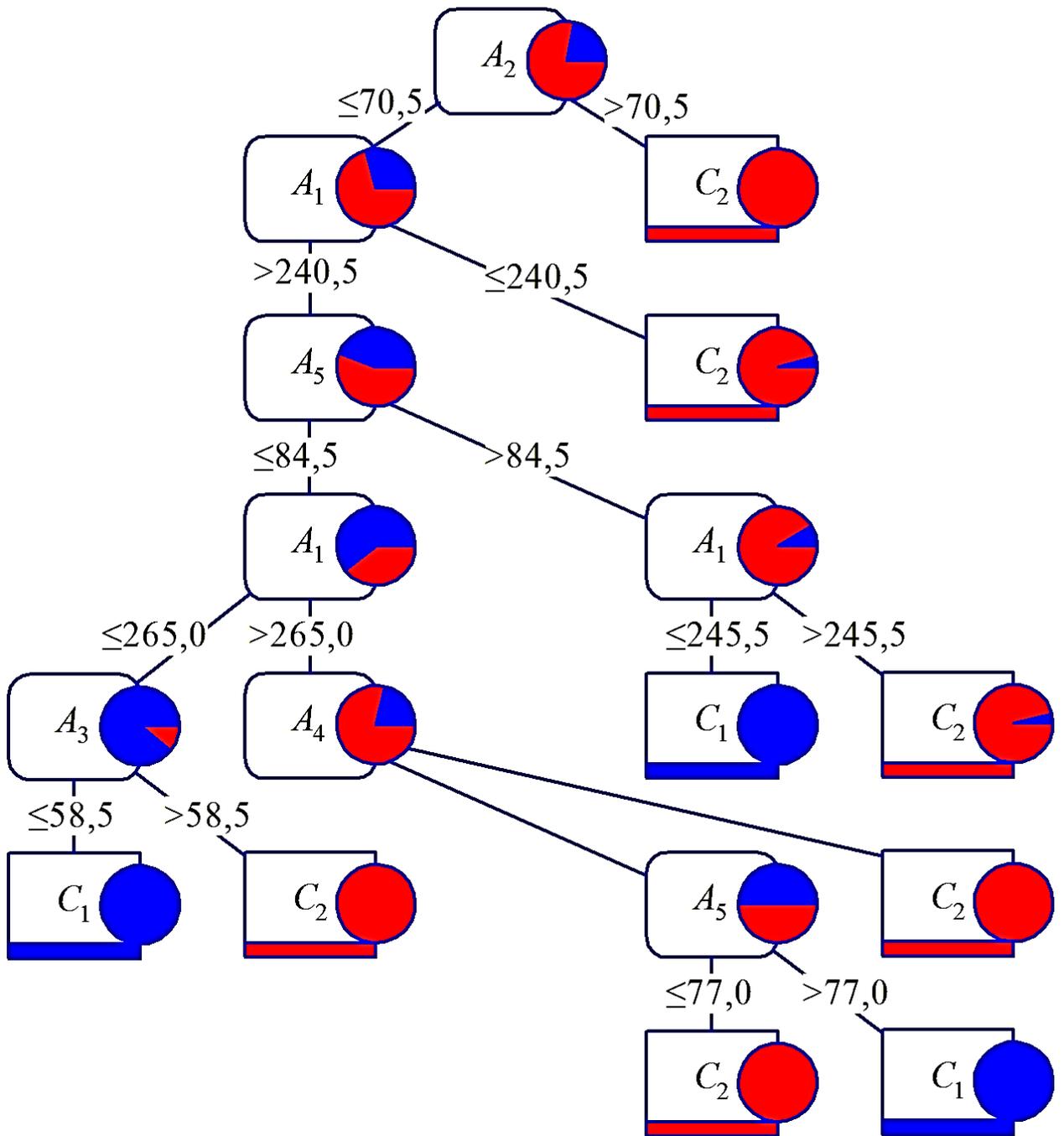


Рисунок 2.5 – Дерево классификации, полученный с помощью алгоритма CART (синим цветом отмечен класс C1 «качественное литье», синим цветом - класс C2 «брак»)

Построенные классификаторы обладают различным количеством листов и узлов:

- структура дерева, полученного с помощью алгоритма ID3, включает в себя 17 узлов и 9 листов;

- структура дерева, полученного с помощью алгоритма C4.5, включает в себя 23 узлов и 12 листов;

- структура дерева, полученного с помощью алгоритма CART, включает в себя 17 узлов и 9 листов.

Наилучшим вариантом является наиболее компактное дерево (с малым количеством узлов и листов), так как в том случае, хранящиеся в дереве правила являются наиболее общими (распространяющимися на большее количество объектов обучающей выборки). Поэтому исключим из рассмотрения дерево, построенное с использованием алгоритма C4.5 (рисунок 2.5).

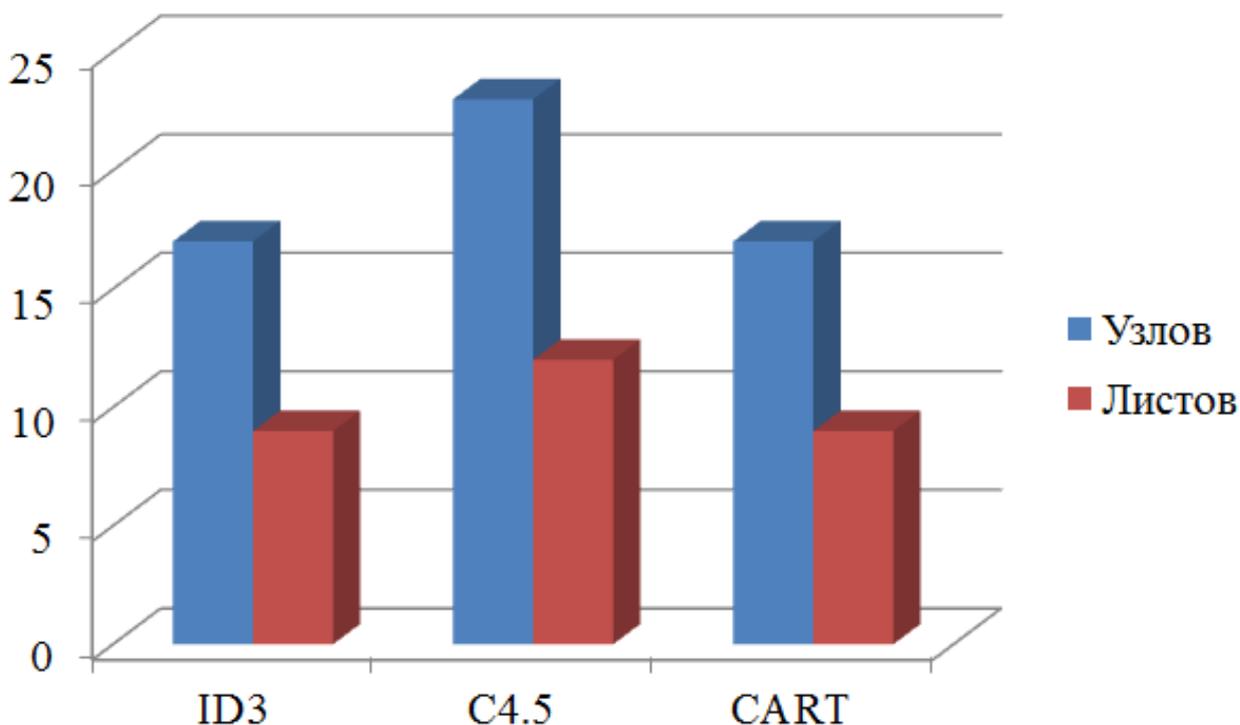


Рисунок 2.5 – Сравнение структуры полученных с помощью разных алгоритмов деревьев принятия решений

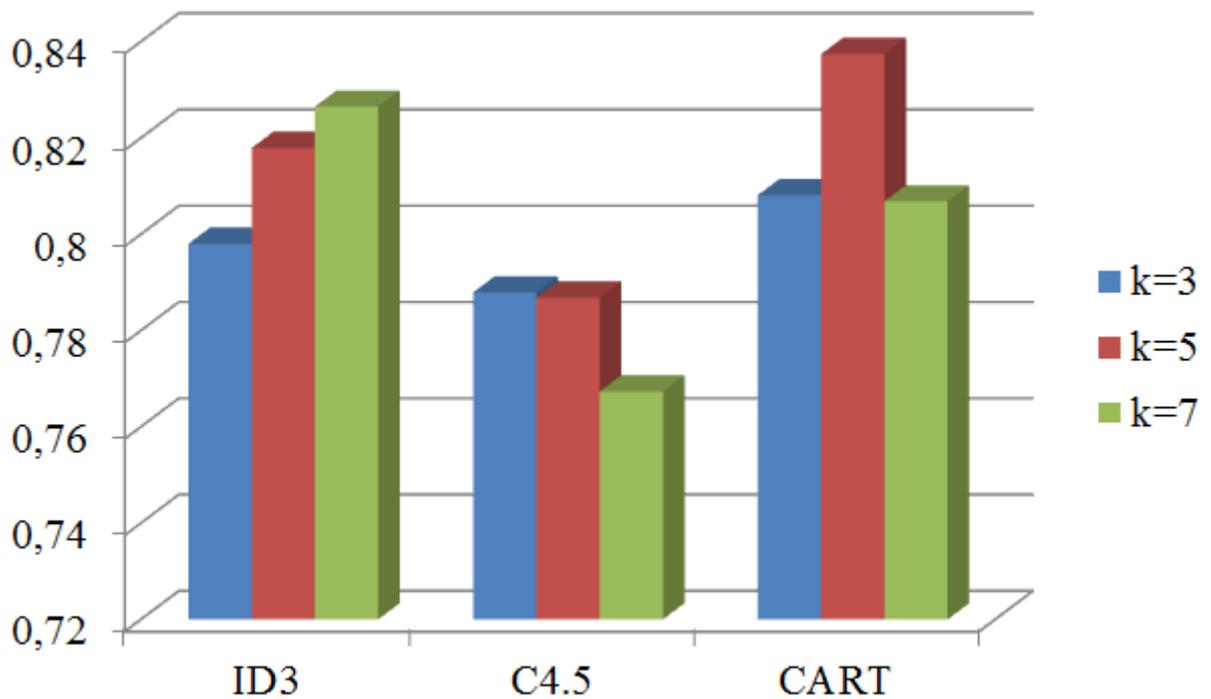


Рисунок 2.5 – Сравнение точности классификаторов методом перекрёстной проверки при различном количестве k частей

Тестирование точности методом перекрёстной проверки при различных значениях параметров k (3; 5 и 7) показал, при использовании алгоритмов ID3 и CART точность находится около значения 80%. Результаты тестирования точности методом перекрёстной проверки показаны на рисунке 2.5.

Для того, что бы сделать окончательный выбор в пользу одного из деревьев была проведена дополнительное тестирование точности методом Leave-One-Out.

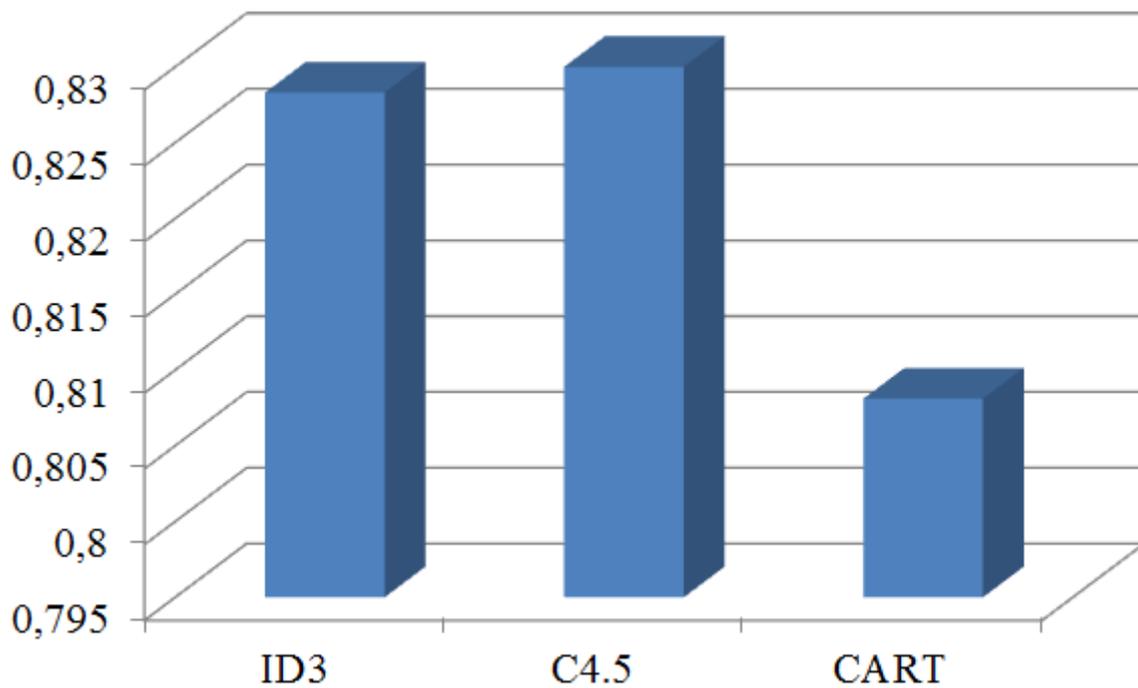


Рисунок 2.5 – Сравнение точности классификаторов с использованием метода Leave-One-Out

Результаты сравнение точности классификаторов с использованием метода Leave-One-Out показаны на рисунке 2.5. Как видно из сравнения точность при использовании алгоритма ID3 выше, чем при использовании алгоритма CART.

Таким образом, лучшим классификатором для диагностики качества деталей выполненных литьем является дерево, представленное на рисунке 2.3, построенное на основе алгоритма ID3.

3 ПРАКТИЧЕСКАЯ РЕАЛИЗАЦИЯ ПРЕДЛОЖЕННЫХ РЕШЕНИЙ

3.1 Описание разработанного программного обеспечения

В ходе выполнения бакалаврской работы на языке программирования Java было разработано программное обеспечение, позволяющее строить деревья принятия решений для диагностики качества деталей, выполненных литьем под давлением. Основные функциональные особенности разработанного программного обеспечения:

- наличие GUI;
- возможность построение деревьев принятия решений с помощью алгоритма ID3;
- графическое представление дерева принятия решений;
- автоматическая выгрузка экспериментальных данных (обучающей выборки) из xlsx файла, находящегося в паке с проектом;
- возможность просмотра и редактирования обучающей выборки с помощью встроенных средств;
- возможность просмотра объектов обучающей выборки, содержащихся в узлах;
- возможность просмотра объектов обучающей выборки, содержащихся в листах дерева.

После запуска программного обеспечения производится автоматическая загрузка обучающей выборки из файла xlsx, содержащегося в папке проекта. Управление процессом построения классификатора осуществляется с помощью верхнего меню. Пункт «новый проект» осуществляет повторную загрузку обучающей выборки из xlsx файла. Пункт «сгенерировать классификатор» отвечает за построение и графическое отображение дерева принятия решений. Пункт «Пошаговое построение» отвечает за поэтапное просмотр построения дерева принятия решений (на каждом этапе формируется по одному узлу), пункт «Обучающая выборка»

открывает окно для просмотра и редактирования обучающей выборки, пункт «Выход» закрывает приложение.

Внешний вид главного окна приложения представлено на рисунке 3.1.

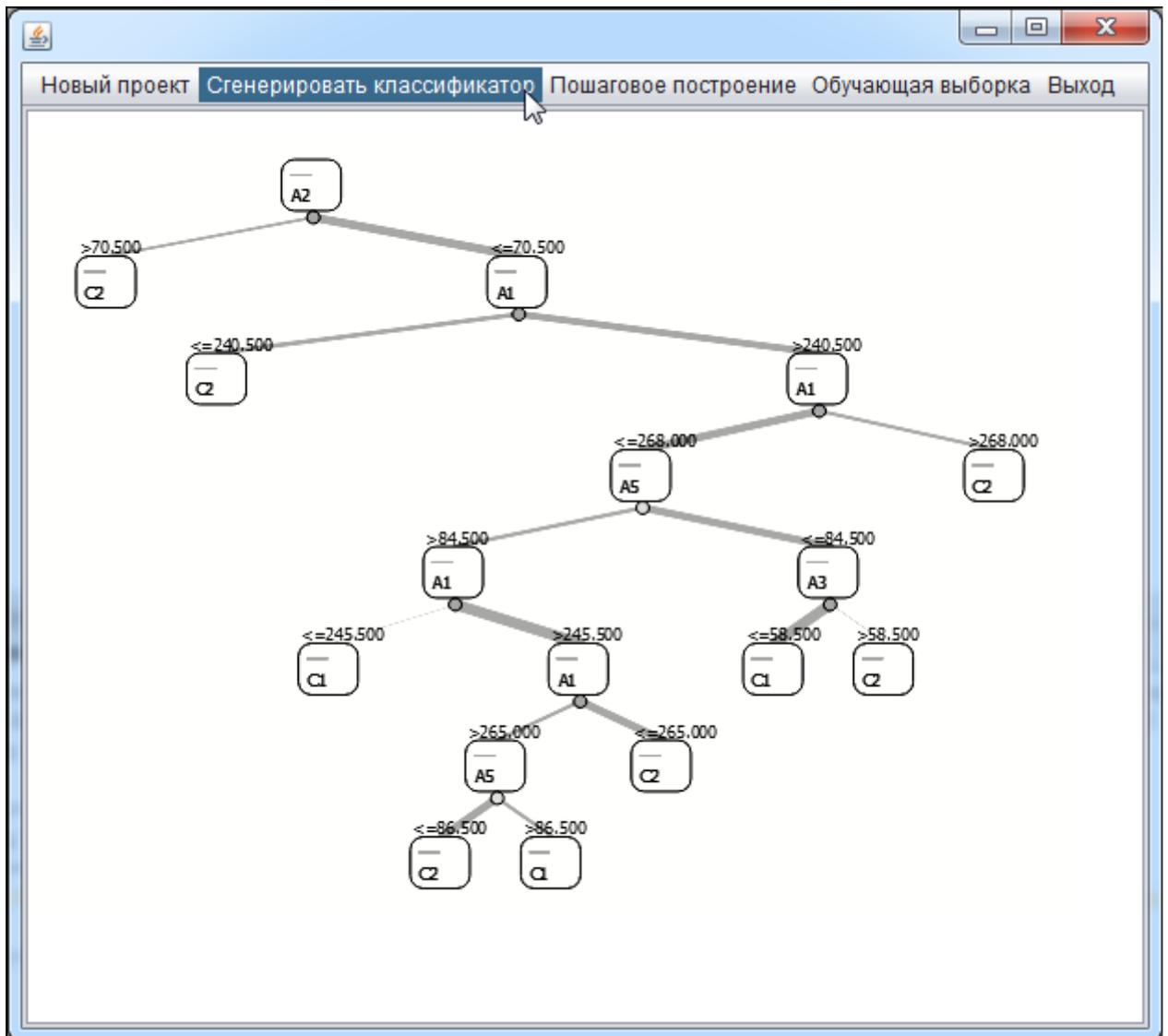


Рисунок 3.1 – Главное окно приложения

3.2 Пример использования классификаторов

Рассмотрим пример использования построенного классификатора качества литья детали на примере случая, когда деталь формируется в следующих условиях:

- температура массы пластика $A_1=267$ °С;
- температура формы $A_2=56$ °С;
- температура зоны загрузки $A_3=37$ °С;
- давление $A_4=83$ бар;
- температура сушки $A_5=84$ °С.

Для определения класса литья (одного из двух вариантов: «С1» –деталь получена без дефектов литья, «С2» –деталь с одним или несколькими дефектами литья) необходимо осуществить спуск по дереву до одного из листов.

Так в корневом узле осуществляется проверка по значению температуры формы A_2 , затем проверка температуры массы пластика A_1 , затем проверка температуры зоны загрузки A_3 .

Как видно из рисунка 3.5 при данном сочетании параметров при осуществлении спуска мы попадаем в лист с классом «С1». Это значит, что с учетом показателей точности классификации, которые мы определяли ранее, с точностью 83% данная деталь не имеет дефектов и является качественной.

Аналогичным образом осуществляется классификация качества деталей, выполненных литьем, при любом другом сочетании параметров A_1 , A_2 , A_3 , A_4 , A_5 .

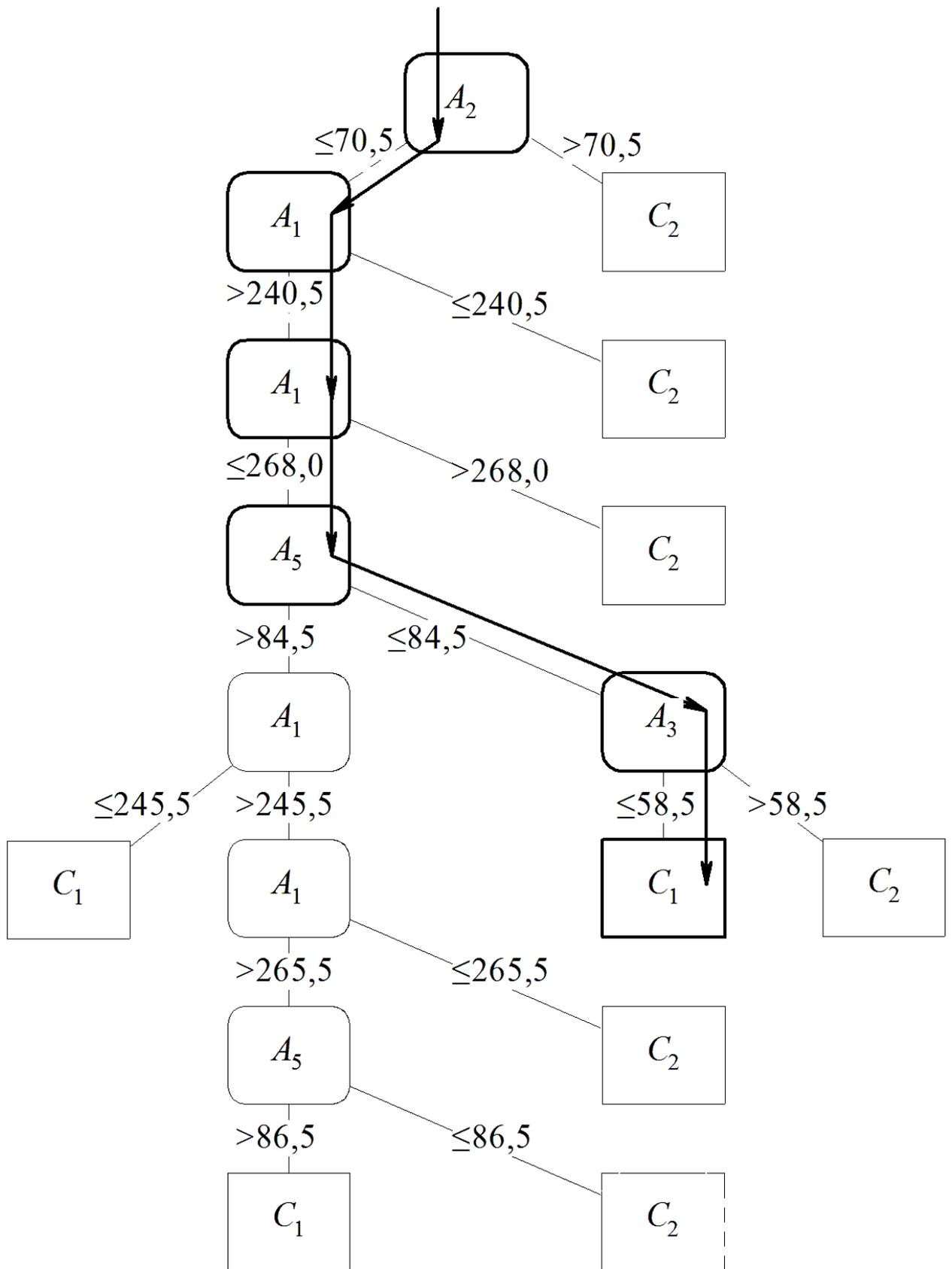


Рисунок 3.2 – Пример работы классификатора для вектора входных параметров x_t

ЗАКЛЮЧЕНИЕ

В ходе выполнения бакалаврской работы были сделаны следующие выводы:

1. Задачу диагностики качества деталей, можно рассматривать как задачу классификации, в которую входит определение зависимости между измеряемыми параметрами и получаемым классом качества детали.

2. Благодаря развитию технологий машинного обучения, построение классификаторов качества можно осуществлять автоматизировано, на основе накопленных производственных экспериментальных данных.

3. На основе данных предоставленных предприятием ЗАО «Мегапласт» по изготовлению методом литья под давлением накладки на левую стойку из АБС пластика для автомобиля Largus с использованием алгоритмов построения деревьев принятия решений (ID3, C4.5, CART) были разработаны классификаторы качества деталей.

4. Была изучена структура полученных классификаторов: дерево, полученное с помощью алгоритма ID3, включает в себя 17 узлов и 9 листов; дерево, полученное с помощью алгоритма C4.5, включает в себя 23 узлов и 12 листов; дерево, полученное с помощью алгоритма CART, включает в себя 17 узлов и 9 листов. Предпочтительным вариантом являются компактные деревья, так в этом случае, хранящиеся в них правила являются более общими (распространяются на большее количество случаев).

5. Была изучена точность полученных классификаторов. Тестирование точности методом перекрёстной проверки при различных значениях параметров k (3; 5 и 7) показал, при использовании алгоритмов ID3 и CART точность классификации находится на уровне 80-85%. Что является удовлетворительным результатом.

6. Для автоматизации построения классификаторов (в виде деревьев принятия решений) было разработано и протестировано программное обеспечение на языке Java.

7. Опубликовано 1 одна статья по теме исследования.

СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ

1. Li, F. Fuzzy ID3 Algorithm Based on Generating Hartley Measure [Text] / Fachao Li, Dandan Jiang // International Conference on Web Information Systems and Mining – International Conference, WISM 2011, Taiyuan, China, September 24-25, 2011. Proceedings, Part II: Web Information Systems and Mining. – Springer-Verlag Berlin Heidelberg 2011. – pp. 188-195
2. Zhu, L. Improvement of Decision Tree ID3 Algorithm [Text] / Lin Zhu, Yang Yang // International Conference on Collaborative Computing: Networking, Applications and Worksharing – 12th International Conference, CollaborateCom 2016, Beijing, China, November 10–11, 2016. Proceedings: Collaborate Computing: Networking, Applications and Worksharing. – ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2017. – pp. 595-600
3. Narazaki, H. A method to use uncertain domain knowledge in the induction of classification knowledge based on ID3 [Text] / Hiroshi Narazaki, Ichiro Shigaki // International Workshop on Fuzzy Logic in Artificial Intelligence – IJCAI'97 Workshop Nagoya, Japan, August 23–24, 1997. Selected and Invited Papers: Fuzzy Logic in Artificial Intelligence. – Springer-Verlag Berlin Heidelberg 1999. – pp. 45-61
4. Auer, J. Investigating ID3-Induced Rules from Low-Dimensional Data Cleaned by Complete Case Analysis [Text] / Jeanette Auer, Richard Hall // Australasian Joint Conference on Artificial Intelligence – 17th Australian Joint Conference on Artificial Intelligence, Cairns, Australia, December 4-6, 2004. Proceedings: AI 2004: Advances in Artificial Intelligence. – Springer-Verlag Berlin Heidelberg 2004. – pp. 414-424
5. García, V. Web Attack Detection Using ID3 [Text] / Víctor H. García, Raúl Monroy, Maricela Quintana // IFIP World Computer Congress, TC 12 – IFIP 19th World Computer Congress, TC 12: Professional Practice Stream, August 21–24, 2006. Santiago, Chile: Professional Practice in

Artificial Intelligence. – International Federation for Information Processing 2006. – pp. 323-332

6. Min, F. A Competition Strategy to Cost-Sensitive Decision Trees [Text] / Fan Min, William Zhu // International Conference on Rough Sets and Knowledge Technology – 7th International Conference, RSKT 2012, Chengdu, China, August 17-20, 2012. Proceedings: Rough Sets and Knowledge Technology. – Springer-Verlag Berlin Heidelberg 2012. – pp. 359-368

7. Chang, J. Genetic Algorithm Based Fuzzy ID3 Algorithm [Text] / Jyh-Yeong Chang, Chien-Wen Cho, Su-Hwang Hsieh, Shi-Tsung Chen // International Conference on Neural Information Processing – 11th International Conference, ICONIP 2004, Calcutta, India, November 22-25, 2004. Proceedings: Neural Information Processing. – Springer-Verlag Berlin Heidelberg 2004. – pp. 989-995

8. Bartczuk, Ł. A New Version of the Fuzzy-ID3 Algorithm [Text] / Łukasz Bartczuk, Danuta Rutkowska // International Conference on Artificial Intelligence and Soft Computing – 8th International Conference, Zakopane, Poland, June 25-29, 2006. Proceedings: Artificial Intelligence and Soft Computing – ICAISC 2006. – Springer-Verlag Berlin Heidelberg 2006. – pp. 1060-1070

9. Zheng, Z. Scaling up the rule generation of C4.5 [Text] / Zijian Zheng // Pacific-Asia Conference on Knowledge Discovery and Data Mining – Second Pacific-Asia Conference, PAKDD-98 Melbourne, Australia, April 15–17, 1998. Proceedings: Research and Development in Knowledge Discovery and Data Mining. – Springer-Verlag Berlin Heidelberg 1998. – pp. 348-359

10. Jiang, S. A Combined Classification Algorithm Based on C4.5 and NB [Text] / ShengYi Jiang, Wen Yu // International Symposium on Intelligence Computation and Applications – Third International Symposium, ISICA 2008 Wuhan, China, December 19-21, 2008. Proceedings: Advances in Computation and Intelligence. – Springer-Verlag Berlin Heidelberg 2008. – pp. 350-359

11. Jia, D. Publication Recommendations of Manuscripts Using Improved C4.5 Decision Tree Algorithm [Text] / Didi Jia, Wu Xie, Zhiyong Chen, Baohua Qiang // International Conference on Intelligent and Interactive Systems and Applications – Proceedings of the 2nd International Conference on Intelligent and Interactive Systems and Applications (IISA2017): Advances in Intelligent Systems and Interactive Applications. – Springer International Publishing AG 2018. – pp. 710-714

12. Jiang, S. A Combination Classification Algorithm Based on Outlier Detection and C4.5 [Text] / ShengYi Jiang, Wen Yu // International Conference on Advanced Data Mining and Applications – 5th International Conference, ADMA 2009, Beijing, China, August 17-19, 2009. Proceedings: Advanced Data Mining and Applications. – Springer-Verlag Berlin Heidelberg 2009. –pp. 504-511

13. Karimi, K. Logical Decision Rules: Teaching C4.5 to Speak Prolog [Text] / Kamran Karimi, Howard J. Hamilton – International Conference on Intelligent Data Engineering and Automated Learning // Second International Conference Shatin, N.T., Hong Kong, China, December 13–15, 2000. Proceedings: Intelligent Data Engineering and Automated Learning — IDEAL 2000. Data Mining, Financial Engineering, and Intelligent Agents. – Springer-Verlag Berlin Heidelberg 2000. – pp. 85-90

14. Zhang, L. C4.5 or Naive Bayes: A Discriminative Model Selection Approach [Text] / Lungan Zhang, Liangxiao Jiang, Chaoqun Li // International Conference on Artificial Neural Networks – 25th International Conference on Artificial Neural Networks, Barcelona, Spain, September 6-9, 2016. Proceedings, Part I: Artificial Neural Networks and Machine Learning – ICANN 2016. – Springer International Publishing Switzerland 2016. – pp. 419-426

15. Karimi, K. Finding Temporal Relations: Causal Bayesian Networks vs. C4.5 [Text] / Kamran Karimi, Howard J. Hamilton // International Symposium on Methodologies for Intelligent Systems – 12th International

Symposium, ISMIS 2000 Charlotte, NC, USA, October 11–14, 2000. Proceedings: Foundations of Intelligent Systems. – Springer-Verlag Berlin Heidelberg 2000. – pp. 266-273

16. Mahmood, A. An Improved CART Decision Tree for Datasets with Irrelevant Feature [Text] / Ali Mirza Mahmood, Mohammad Imran, Naganjaneyulu Satuluri, Mrithyumjaya Rao Kuppa, Vemulakonda Rajesh // International Conference on Swarm, Evolutionary, and Memetic Computing – Second International Conference, SEMCCO 2011, Visakhapatnam, Andhra Pradesh, India, December 19-21, 2011. Proceedings, Part I: Swarm, Evolutionary, and Memetic Computing. – Springer-Verlag Berlin Heidelberg 2011. – pp. 539-549

17. Povalej, P. Verifying Clinical Criteria for Parkinsonian Disorders with CART Decision Trees [Text] / Petra Povalej, Gregor Štiglic, Peter Kokol, Bruno Stiglic, Irene Litvan, Dušan Flisar // International Conference on Knowledge-Based and Intelligent Information and Engineering Systems – 8th International Conference, KES 2004, Wellington, New Zealand, September 20-25, 2004. Proceedings, Part II: Knowledge-Based Intelligent Information and Engineering Systems. – Springer-Verlag Berlin Heidelberg 2004. – pp. 1018-1024

18. Pal, U. A Study on CART Based on Maximum Probabilistic-Based Rough Set [Text] / Utpal Pal, Sharmistha Bhattacharya (Halder), Kalyani Debnath // International Conference on Mining Intelligence and Knowledge Exploration – 5th International Conference, MIKE 2017, Hyderabad, India, December 13–15, 2017. Proceedings: Mining Intelligence and Knowledge Exploration. – Springer International Publishing AG 2017. – pp. 412-423

19. Seera, M. A Hybrid FMM-CART Model for Fault Detection and Diagnosis of Induction Motors [Text] / Manjeevan Seera, CheePeng Lim, Dahaman Ishak // International Conference on Neural Information Processing – 18th International Conference, ICONIP 2011, Shanghai, China, November 13-

17, 2011. Proceedings, Part III: Neural Information Processing. – Springer-Verlag Berlin Heidelberg 2011. – pp. 730-736

20. Thomas, P. CART for Supply Chain Simulation Models Reduction [Text] / Philippe Thomas, Marie-Christine Suhner, André Thomas // IFIP International Conference on Advances in Production Management Systems – IFIP WG 5.7 International Conference, APMS 2014, Ajaccio, France, September 20-24, 2014. Proceedings, Part III: Advances in Production Management Systems. Innovative and Knowledge-Based Production Management in a Global-Local World. – IFIP International Federation for Information Processing 2014. – pp. 530-537