



# **Сравнение методов машинного обучения для обнаружения дефектов программного обеспечения**

Студент: Климов Илья Сергеевич, ИУ7-72Б

Научный руководитель: Вишневская Татьяна Ивановна

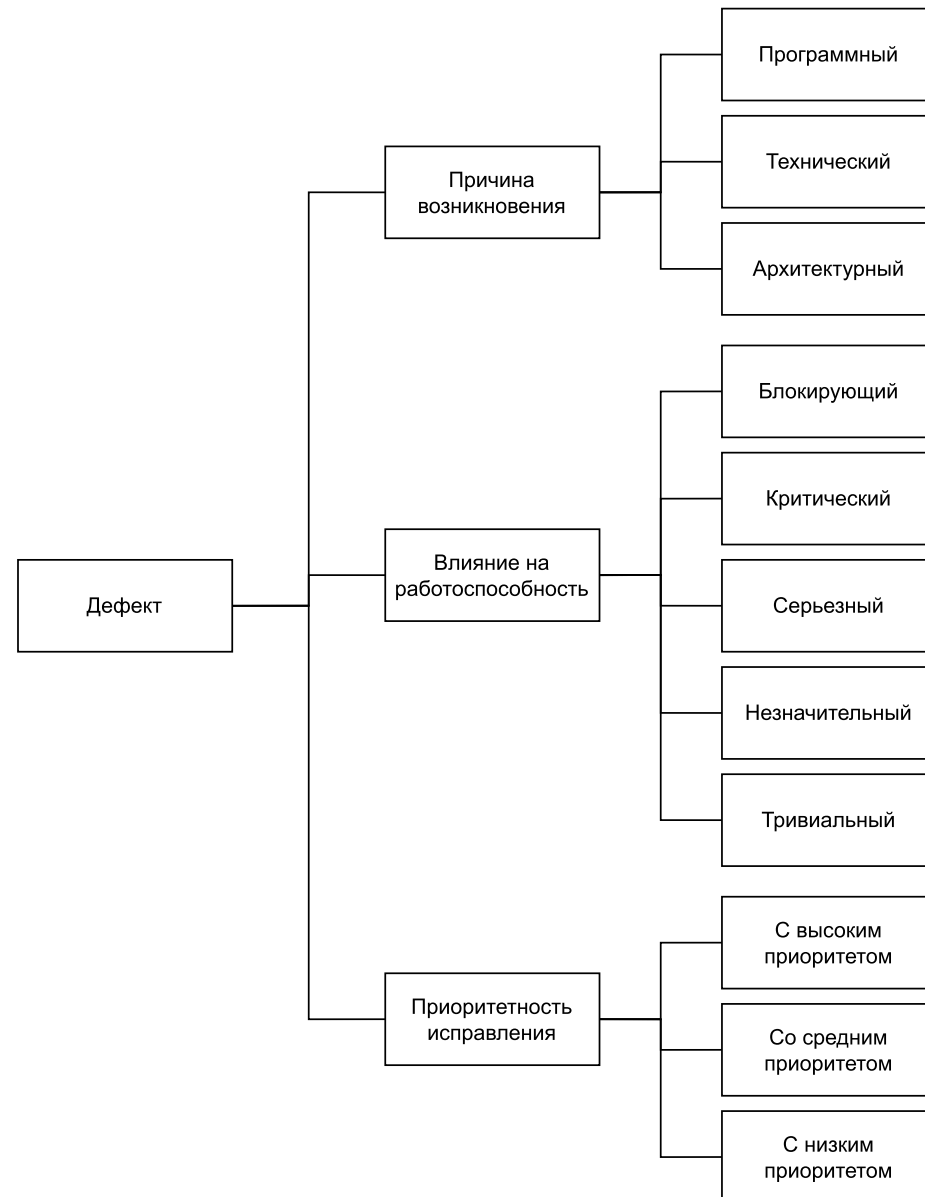
# Цель и задачи

**Цель:** обзор и сравнение методов машинного обучения для обнаружения дефектов ПО.

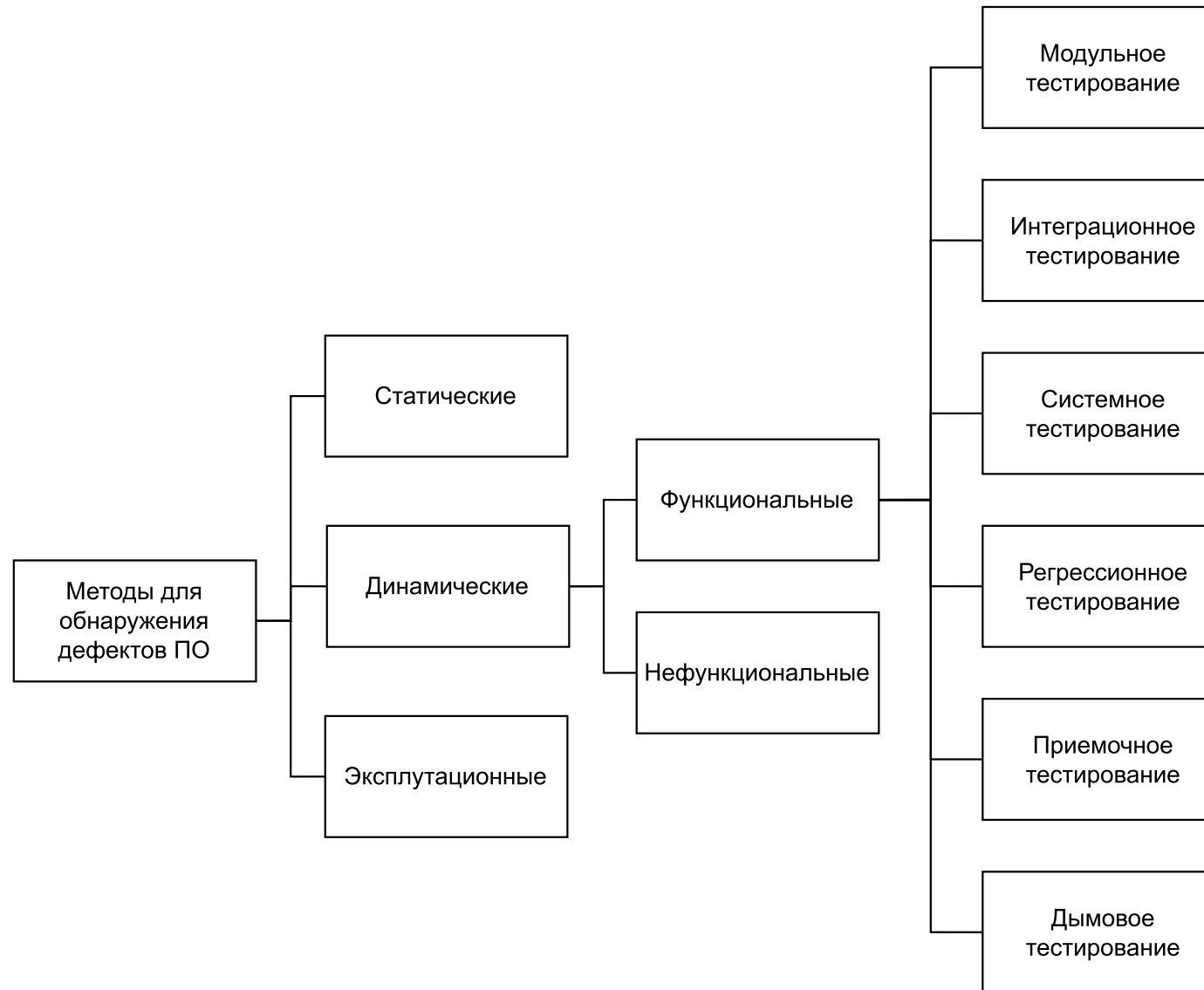
**Задачи:**

- 1) представить обзор дефектов разрабатываемого ПО;
- 2) классифицировать методы для обнаружения дефектов ПО;
- 3) рассмотреть возможность использования методов машинного обучения в данной сфере;
- 4) сформулировать параметры сравнения методов машинного обучения для обнаружения дефектов ПО;
- 5) провести обзор и сравнение существующих методов машинного обучения для обнаружения дефектов ПО;
- 6) формализовать задачу обнаружения дефектов ПО в виде диаграммы в нотации IDEF0.

# Дефекты разрабатываемого ПО



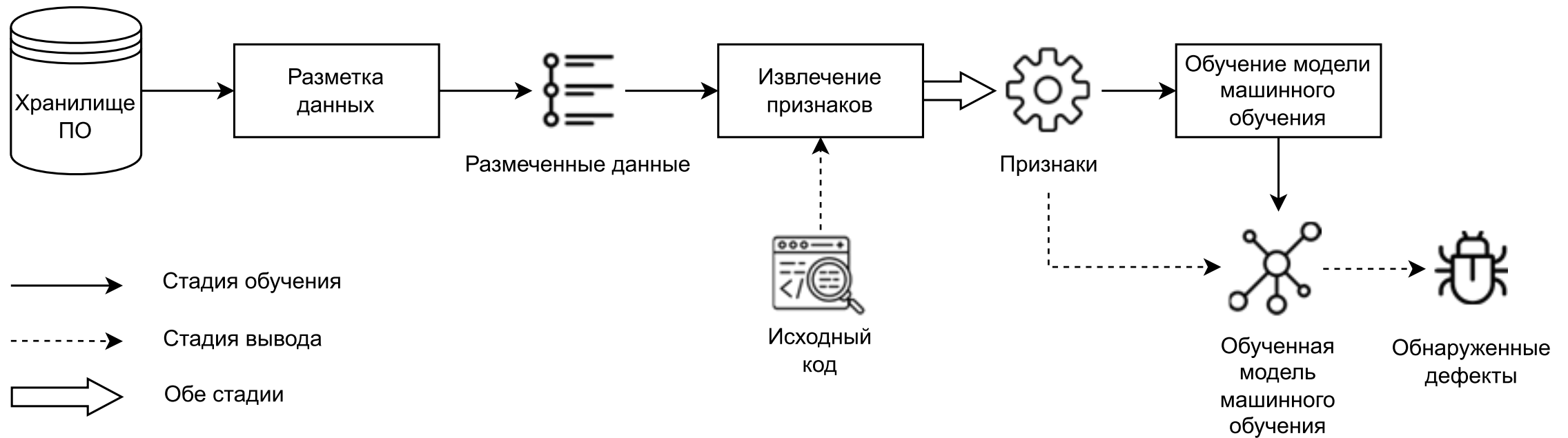
# Методы для обнаружения дефектов ПО



# Методы машинного обучения для обнаружения дефектов ПО

1. Наивный байесовский классификатор.
2. Метод опорных векторов.
3. Дерево решений.
4. Алгоритм случайного леса.
5. Бустинг:
  - адаптивный;
  - градиентный.

# Процесс обучения модели обнаружения дефектов ПО



# Метрики для сравнения алгоритмов

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{F - measure} = \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

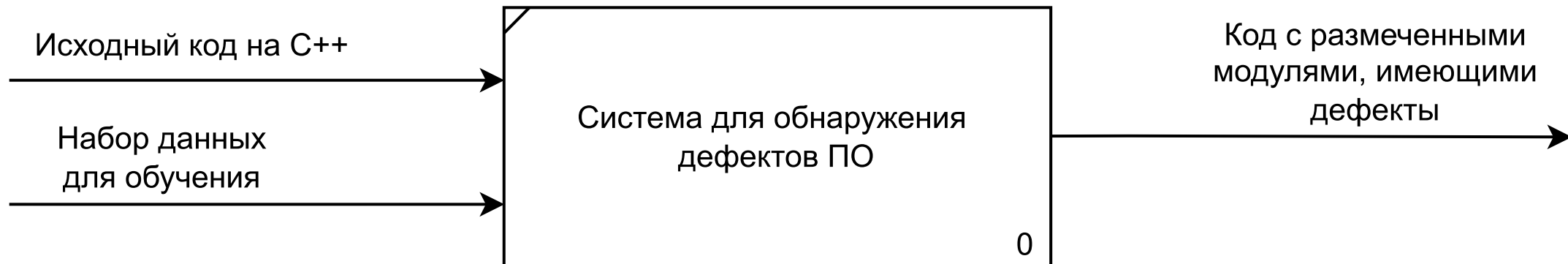
TP – объекты, которые были верно классифицированы как положительные;  
TN – объекты, которые были верно классифицированы как отрицательные;  
FP – объекты, которые были ложно классифицированы как положительные;  
FN – объекты, которые были ложно классифицированы как отрицательные.

# Сравнительная таблица результатов работы алгоритмов

Метрики Алгоритмы	Accuracy (точность)	Precision (точность)	Recall (полнота)	F-measure (F-мера)
Наивный байесовский классификатор	0.795	0.845	0.803	0.849
Метод опорных векторов	0.841	0.901	0.879	0.902
Дерево решений	0.823	0.845	0.878	0.889
Алгоритм случайного леса	0.847	0.903	0.883	0.903
Градиентный бустинг	0.845	0.859	0.863	0.890
Адаптивный бустинг	0.835	0.858	0.861	0.889



# Формализованная постановка задачи



# Заключение

В результате выполнения работы достигнута цель и решены все поставленные задачи:

- 1) представлен обзор дефектов разрабатываемого ПО;
- 2) классифицированы методы для обнаружения дефектов ПО;
- 3) рассмотрена возможность использования методов машинного обучения в данной сфере;
- 4) сформулированы параметры сравнения методов машинного обучения для обнаружения дефектов ПО;
- 5) проведен обзор и сравнение существующих методов машинного обучения для обнаружения дефектов ПО;
- 6) формализована задача обнаружения дефектов ПО в виде диаграммы в нотации IDEF0.