

1. Обработайте набор данных, указанный во втором столбце таблицы 4.1, подготовив его к решению задачи классификации. Выделите целевой признак, указанный в последнем столбце таблицы, и удалите его из данных, на основе которых будет обучаться классификатор. Разделите набор данных на тестовую и обучающую выборку. Постройте классификатор типа, указанного в третьем столбце, для задачи классификации по параметру, указанному в последнем столбце. Оцените точность построенного классификатора с помощью метрик precision, recall и F1 на тестовой выборке.
2. Постройте классификатор типа Случайный Лес (Random Forest) для решения той же задачи классификации. Оцените его качество с помощью метрик precision, recall и F1 на тестовой выборке. Какой из классификаторов оказывается лучше?

Таблица 4.1. Типы классификаторов, наборы данных, признак для классификации

Вариант	Набор данных	Тип классификатора	Классификация по столбцу
1.	HR Analytics: Job Change of Data Scientists https://www.kaggle.com/arashnic/hr-analytics-job-change-of-data-scientists	DecisionTreeClassifier (решающее дерево)	Education level (Graduate – класс 0, остальные уровни – класс 1)
2.	HR Analytics: Job Change of Data Scientists https://www.kaggle.com/arashnic/hr-analytics-job-change-of-data-scientists	LogisticRegression (логистическая регрессия)	Education level (Masters – класс 0, остальные уровни – класс 1)
3.	HR Analytics: Job Change of Data Scientists https://www.kaggle.com/arashnic/hr-analytics-job-change-of-data-scientists	SVM (метод опорных векторов)	Пол gender
4.	Video game sales https://www.kaggle.com/gregorut/videogamesales	DecisionTreeClassifier (решающее дерево)	Platform (DS – класс 0, остальные уровни – класс 1)
5.	Video game sales https://www.kaggle.com/gregorut/videogamesales	LogisticRegression (логистическая регрессия)	Platform (PS2 – класс 0, остальные уровни – класс 1)
6.	Video game sales https://www.kaggle.com/gregorut/videogamesales	SVM (метод опорных векторов)	Genre (Action – класс 0, остальные уровни – класс 1)
7.	Video game sales https://www.kaggle.com/gregorut/videogamesales	DecisionTreeClassifier (решающее дерево)	Genre (Sports – класс 0, остальные уровни – класс 1)
8.	Drug classification https://www.kaggle.com/prathamtripathi/drug-	DecisionTreeClassifier (решающее дерево)	Drug Type (DrugY – класс 0, остальные уровни – класс 1)

	classification		
9.	Drug classification https://www.kaggle.com/prathamtripathi/drug-classification	LogisticRegression (логистическая регрессия)	Drug Type (DrugX – класс 0, остальные уровни – класс 1)
10.	Drug classification https://www.kaggle.com/prathamtripathi/drug-classification	SVM (метод опорных векторов)	Drug Type (DrugA – класс 0, остальные уровни – класс 1)
11.	Students performance in exams https://www.kaggle.com/spscientist/students-performance-in-exams	DecisionTreeClassifier (решающее дерево)	Пол gender
12.	Students performance in exams https://www.kaggle.com/spscientist/students-performance-in-exams	LogisticRegression (логистическая регрессия)	Race ethnicity (group C – класс 0, остальные уровни – класс 1)
13.	Students performance in exams https://www.kaggle.com/spscientist/students-performance-in-exams	SVM (метод опорных векторов)	Race ethnicity (group D – класс 0, остальные уровни – класс 1)
14.	Students performance in exams https://www.kaggle.com/spscientist/students-performance-in-exams	DecisionTreeClassifier (решающее дерево)	Test preparation course
15.	Students performance in exams https://www.kaggle.com/spscientist/students-performance-in-exams	LogisticRegression (логистическая регрессия)	Math Score (выше среднего значения – класс 0, ниже или совпадает – класс 1)
16.	Students performance in exams https://www.kaggle.com/spscientist/students-performance-in-exams	SVM (метод опорных векторов)	Writing Score (выше среднего значения – класс 0, ниже или совпадает – класс 1)
17.	Students performance in exams https://www.kaggle.com/spscientist/students-performance-in-exams	DecisionTreeClassifier (решающее дерево)	Reading Score (выше среднего значения – класс 0, ниже или совпадает – класс 1)
18.	Credit card customers https://www.kaggle.com/sakshigoyal7/credit-card-customers	DecisionTreeClassifier (решающее дерево)	Customer Age (выше среднего значения – класс 0, ниже или совпадает – класс 1)
19.	Credit card customers https://www.kaggle.com/sakshigoyal7/credit-card-	LogisticRegression (логистическая регрессия)	Пол gender

	customers		
20.	Credit card customers https://www.kaggle.com/sakshigoyal7/credit-card-customers	SVM (метод опорных векторов)	Marital Status (married – класс 0, остальные уровни – класс 1)
21.	Credit card customers https://www.kaggle.com/sakshigoyal7/credit-card-customers	DecisionTreeClassifier (решающее дерево)	Marital Status (single – класс 0, остальные уровни – класс 1)
22.	Credit card customers https://www.kaggle.com/sakshigoyal7/credit-card-customers	LogisticRegression (логистическая регрессия)	Education level (graduate и выше – класс 0, ниже – класс 1)
23.	Credit card customers https://www.kaggle.com/sakshigoyal7/credit-card-customers	SVM (метод опорных векторов)	Total relationship count (4 и выше – класс 0, ниже – класс 1)
24.	Credit card customers https://www.kaggle.com/sakshigoyal7/credit-card-customers	DecisionTreeClassifier (решающее дерево)	Credit Limit (выше 10 000 – класс 0, ниже или совпадает – класс 1)
25.	Credit card customers https://www.kaggle.com/sakshigoyal7/credit-card-customers	LogisticRegression (логистическая регрессия)	Total Transaction Amount (ниже или совпадает с 5 000 – класс 0, выше – класс 1)
26.	Student Alcohol Consumption https://www.kaggle.com/uciml/student-alcohol-consumption	DecisionTreeClassifier (решающее дерево)	Final Grade G3 (выше 10 – класс 0, ниже или совпадает с 10 – класс 1)
27.	Student Alcohol Consumption https://www.kaggle.com/uciml/student-alcohol-consumption	LogisticRegression (логистическая регрессия)	Школа Students' school
28.	House Sales in King County https://www.kaggle.com/harlfoxem/housesalesprediction	DecisionTreeClassifier (решающее дерево)	Grade (ниже 7.5 – класс 0, выше или совпадает с 7.5 – класс 1)
29.	House Sales in King County https://www.kaggle.com/harlfoxem/housesalesprediction	LogisticRegression (логистическая регрессия)	Bedrooms (ниже или совпадает с 3 – класс 0, выше – класс 1)
30.	House Sales in King County https://www.kaggle.com/harlfoxem/housesalesprediction	SVM (метод опорных векторов)	Bathrooms (ниже 2.5 – класс 0, выше или совпадает с 2.5 – класс 1)