



САНКТ-ПЕТЕРБУРГСКИЙ
ГОСУДАРСТВЕННЫЙ
ЭКОНОМИЧЕСКИЙ УНИВЕРСИТЕТ



Тема: Анализ и сравнение методов машинного обучения для задачи семантической сегментации дорожной разметки на изображении

Ткаченко Елизавета
Студентка группы Э-2110

Создание датасета



+



Создание датасета

Оригинальное изображение

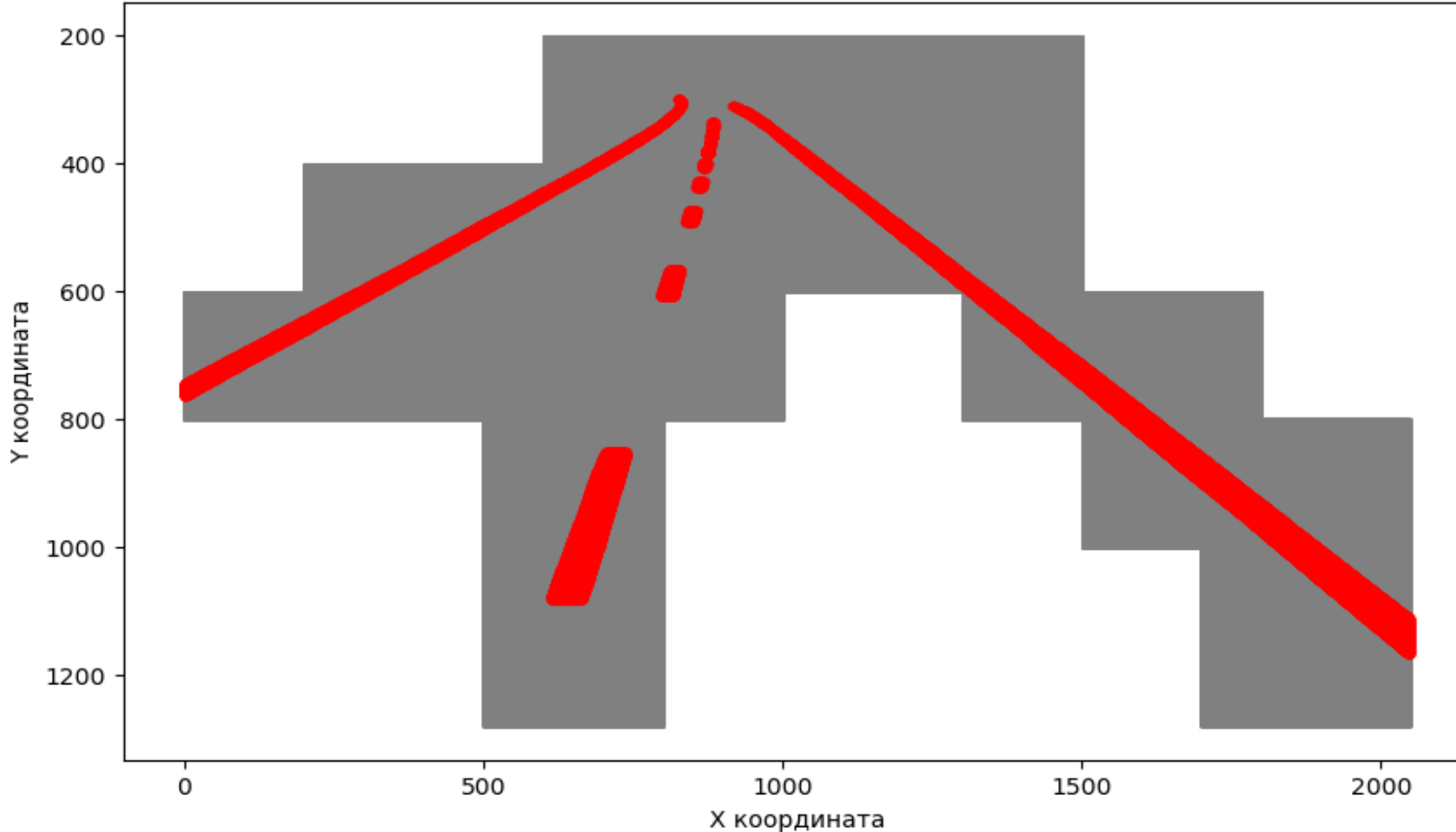


Маска на изображении



Столбцы: **x, y, R, G, B, brightness, class**

Баланс классов



- Класс 1 представлен 2% в наборе данных
- Повышение с 2% до 5% за счет отсеечения частей изображения

Разделение данных и стандартизация

1. На признаки (X) и целевую переменную (y)
2. На тренировочную и тестовую выборки с параметром **stratify=y**

Он который учитывает дисбаланс классов:

Тренировочная выборка		Тестовая выборка	
class		class	
0	94,9461%	0	94,9459%
1	5,0539%	1	5,0541%

3. Стандартизация с помощью StandardScaler

Логистическая регрессия

Лучшие гиперпараметры: 'C': 0.1, 'max_iter': 200, 'solver': 'saga'

Параметр

class_weight='balanced'

Точность = 97,33%

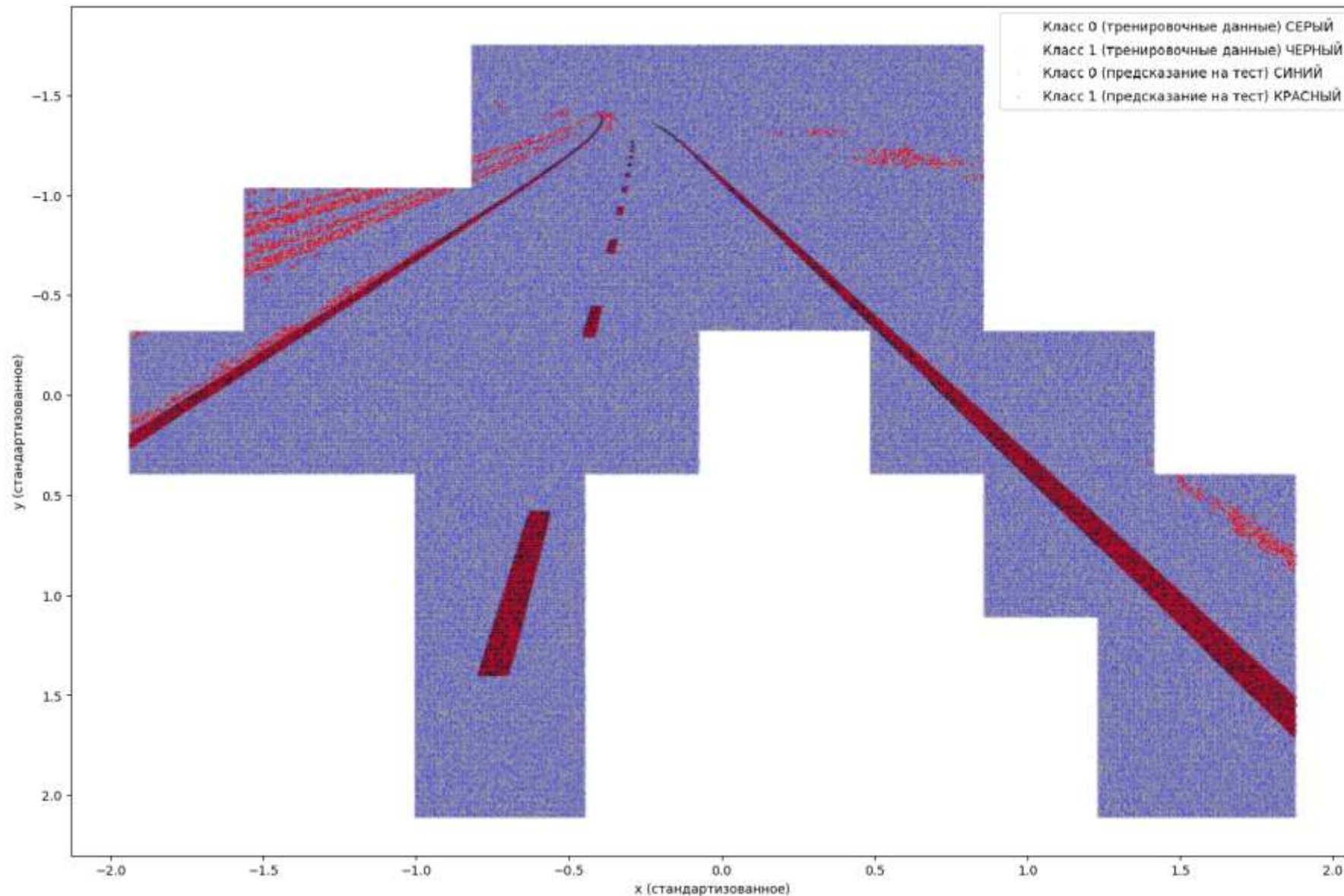
ROC AUC = 99,58%

F1-score для класса 0 = 99%

F1-score для класса 1 = 79%



Логистическая регрессия



Метод опорных векторов

Лучшие гиперпараметры: 'C': 85.09, 'coef0': 0.0, 'degree': 2, 'gamma': 10, 'kernel': 'rbf', 'max_iter': 5000

Параметр

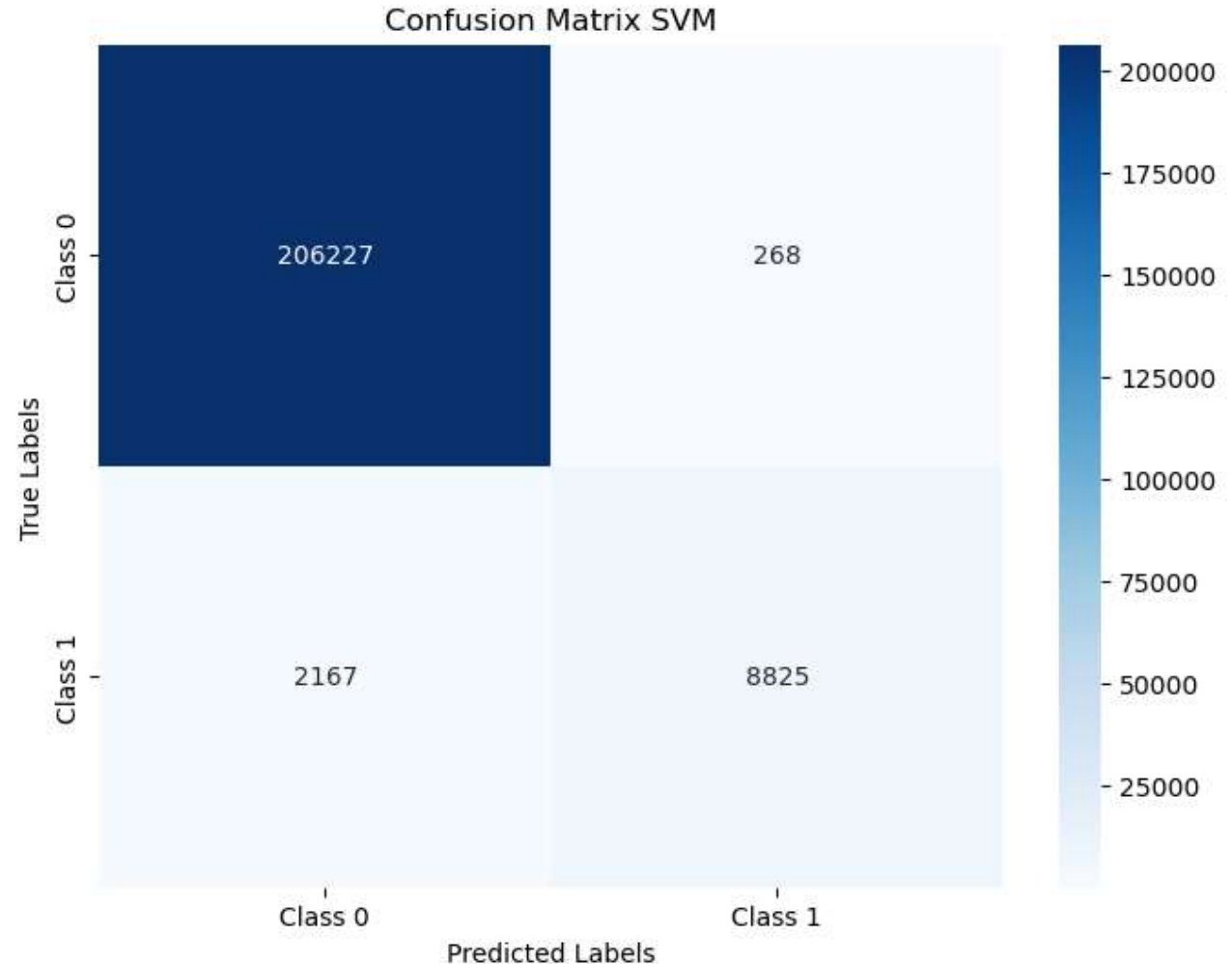
class_weight='balanced'

Точность = 98,88%

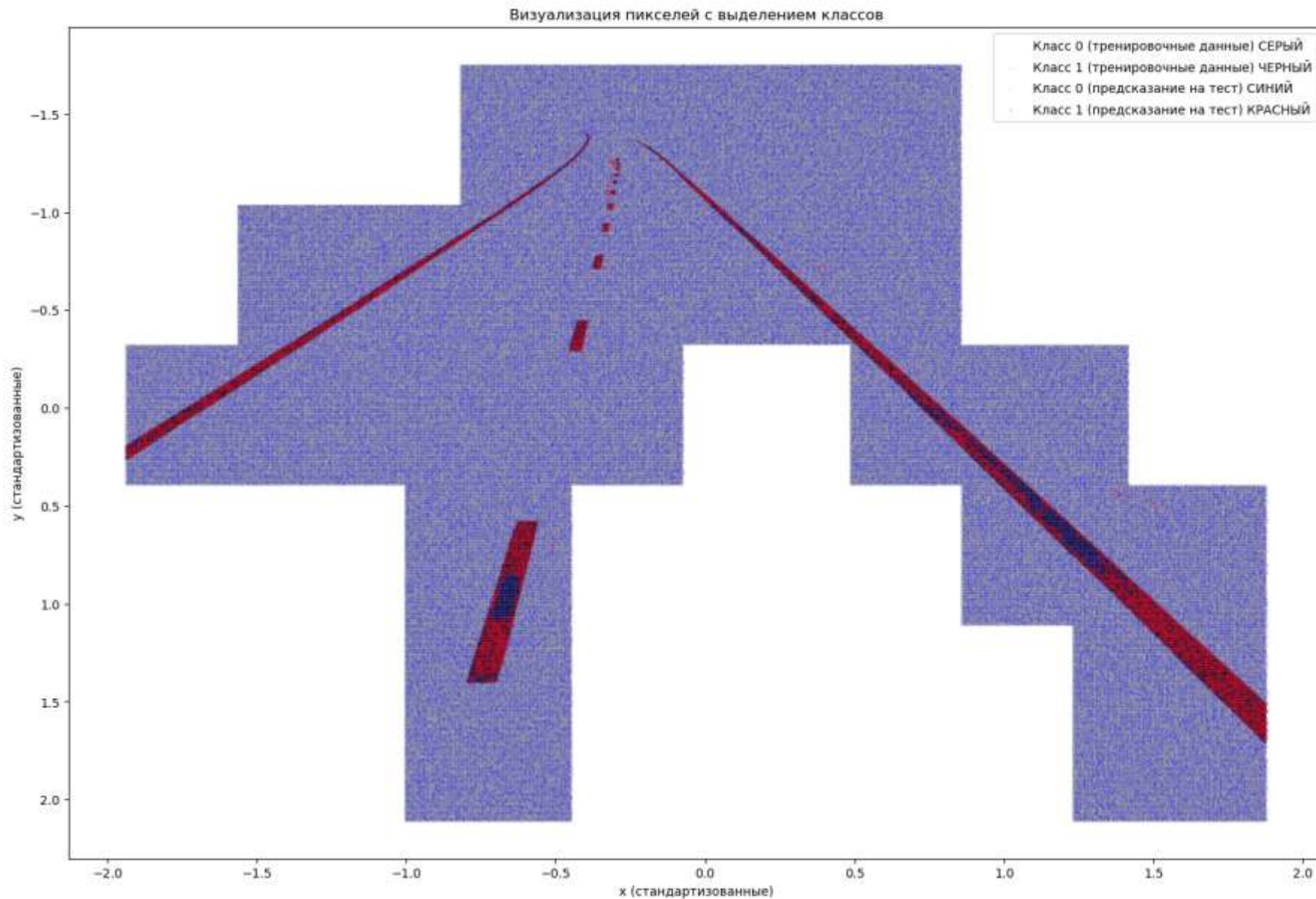
ROC AUC = 99,73%

F1-score для класса 0 = 99%

F1-score для класса 1 = 88%



Метод опорных векторов



Метод k-ближайших соседей

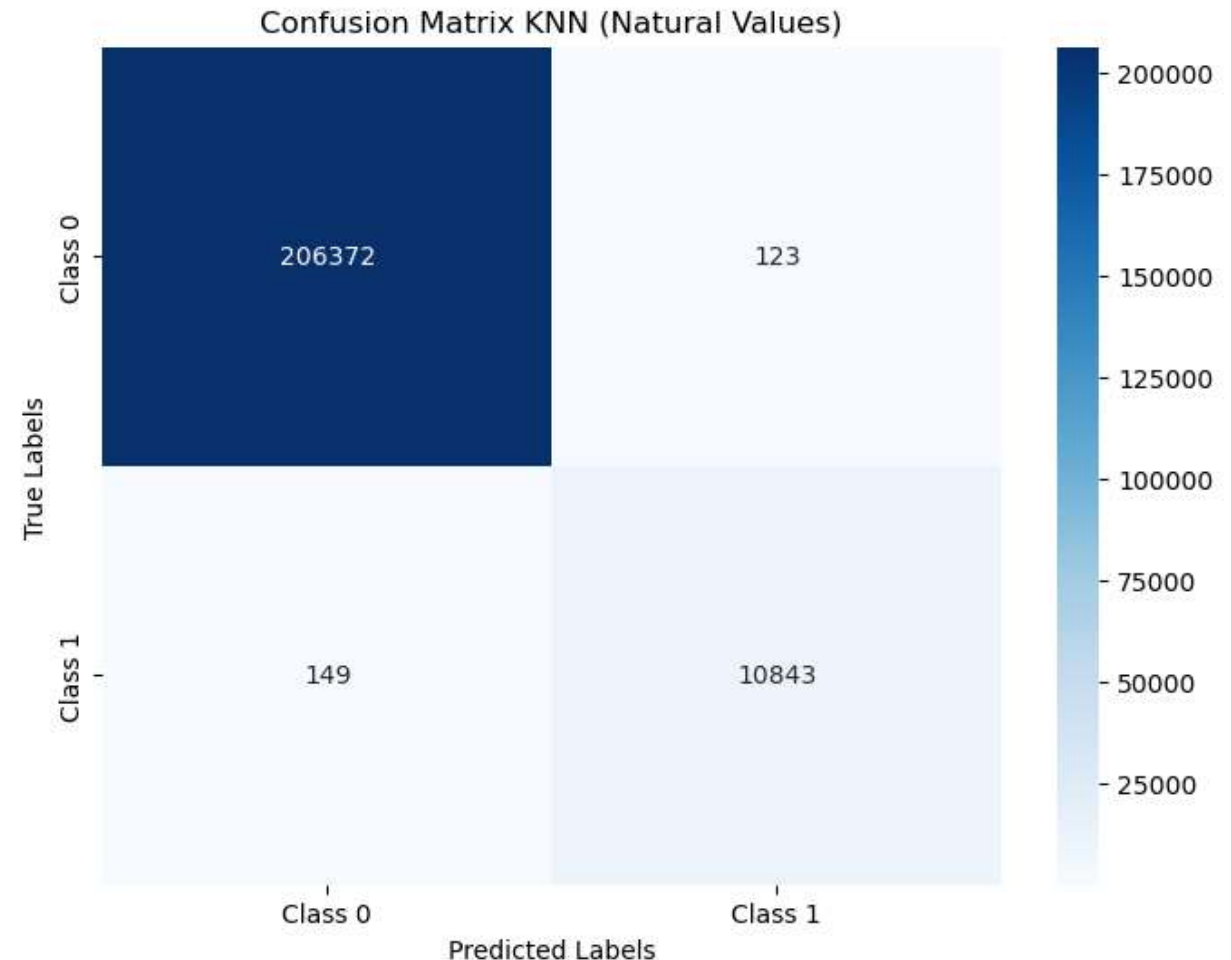
Лучшие гиперпараметры: **'metric': 'manhattan',
'n_neighbors': 5, 'weights': 'distance'**

Точность = 99,87%

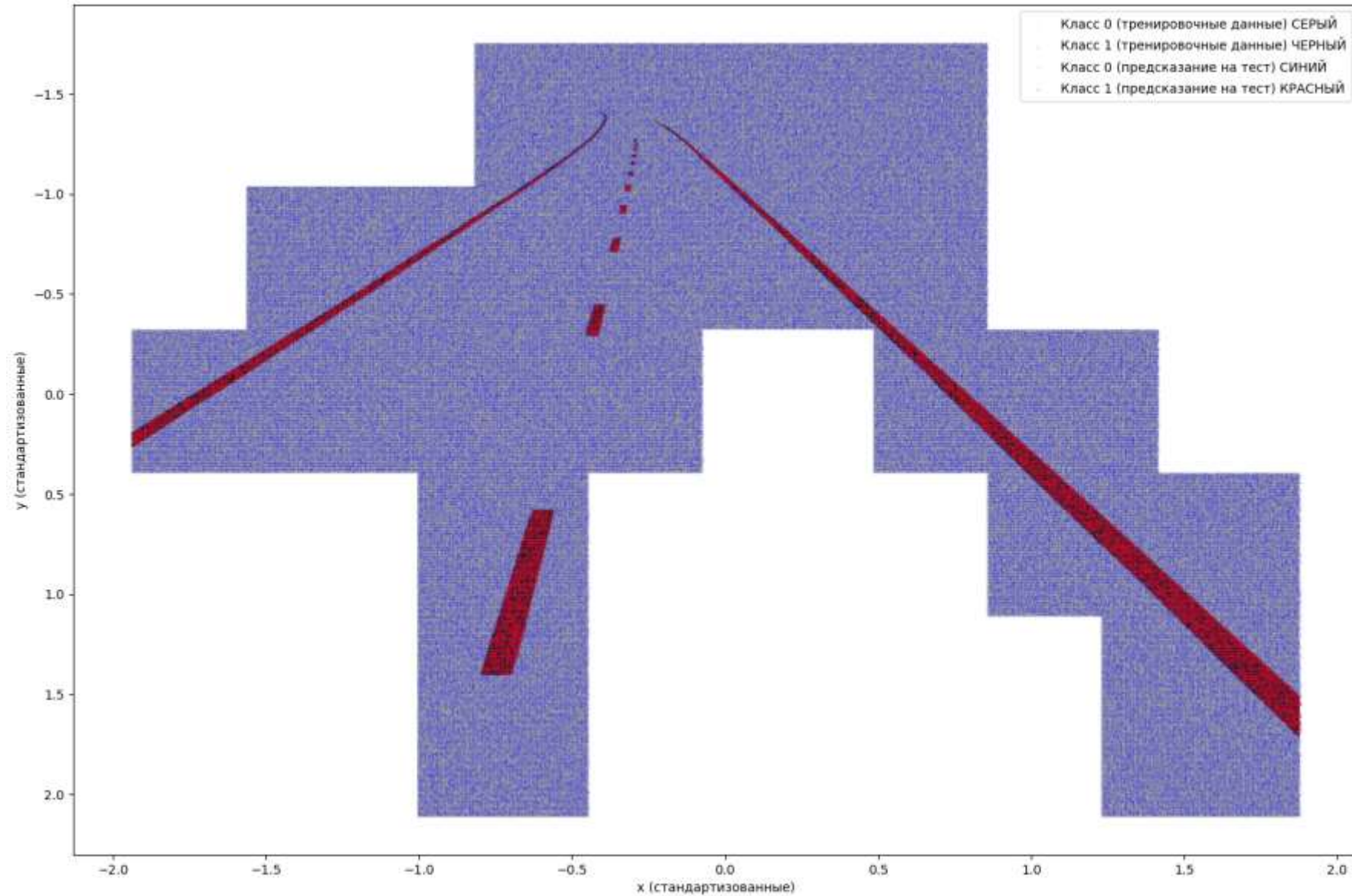
ROC AUC = 99,81%

F1-score для класса 0 = 100%

F1-score для класса 1 = 99%



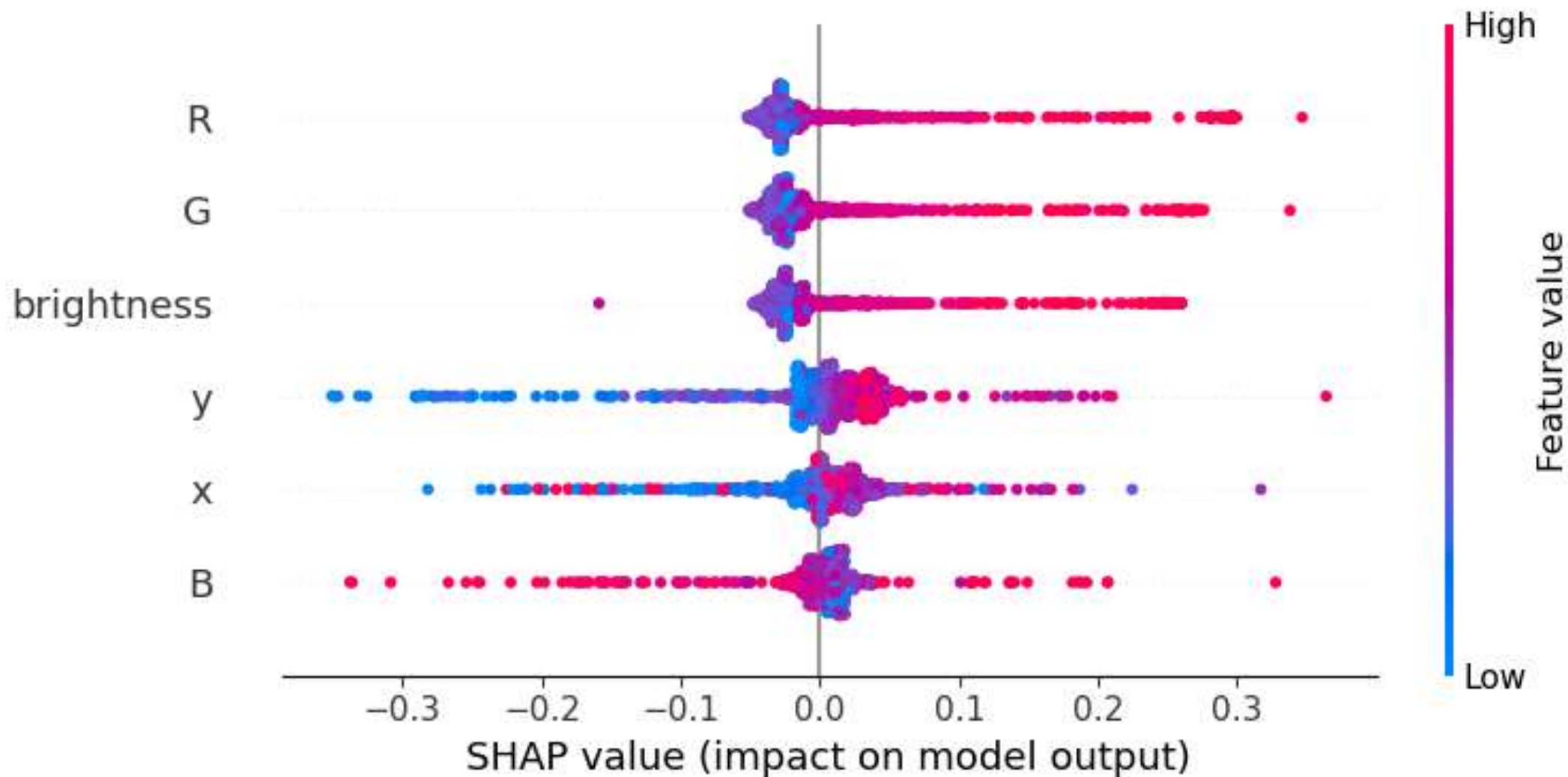
Метод k-ближайших соседей



Сравнительный анализ

Модель	Accuracy	ROC AUC	F1-score для класса 1	Время обучения (сек.)	Время предсказан ия (сек.)
Логистическая регрессия	0,9733	0,9959	0,7900	112,1393	0,0072
SVM	0,9888	0,9973	0,8800	266,8359	162,1469
KNN	0,9987	0,9981	0,9900	3,0566	9,4425

Оценка важности факторов



Модель для полного изображения

Тренировочная выборка = df_1
Изображение №1



Тестовая выборка = df_2
Изображение №2



Метод k-ближайших соседей для полного изображения

Лучшие гиперпараметры:

'metric': 'manhattan',

'n_neighbors': 11,

'weights': 'uniform'

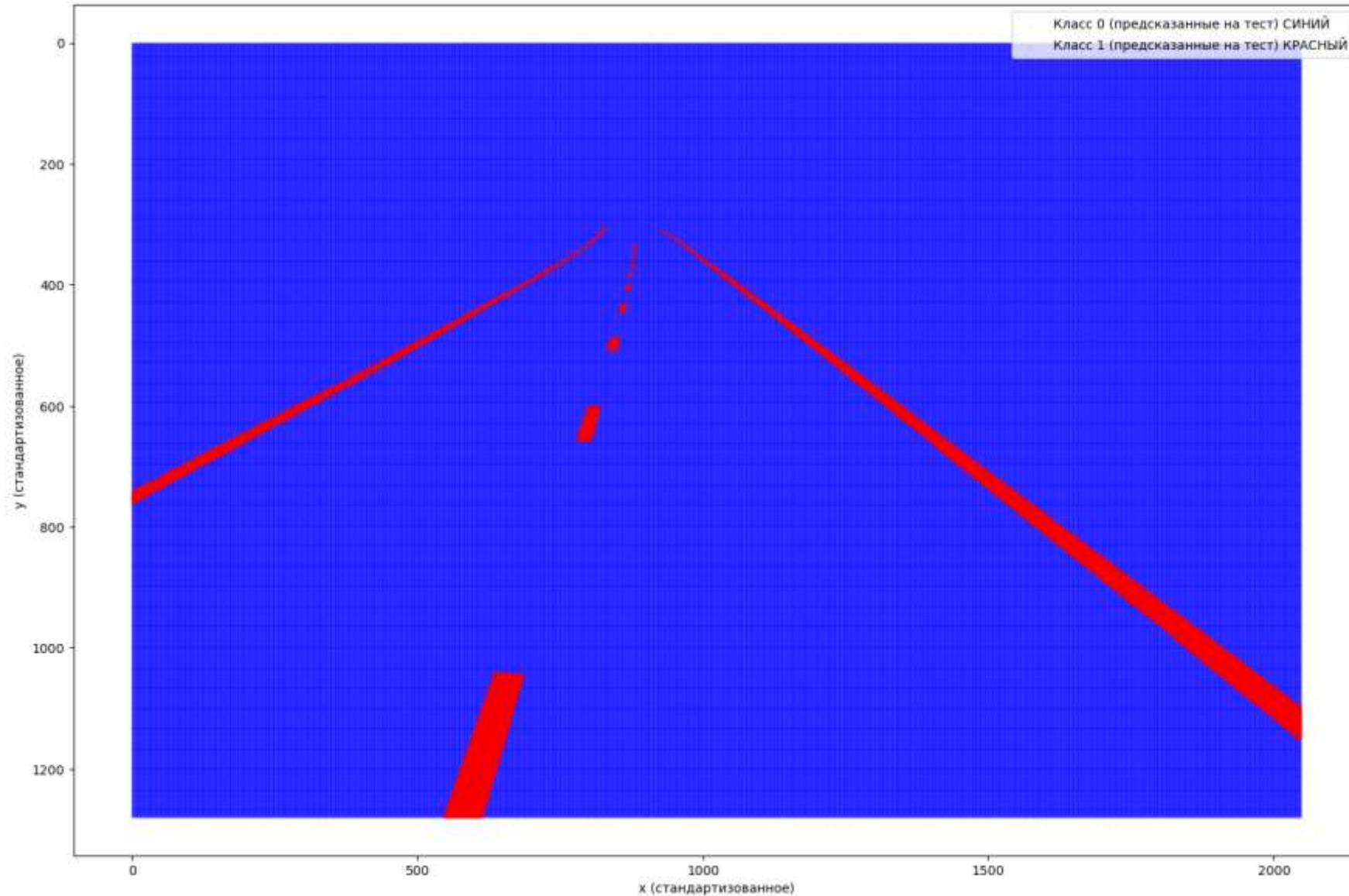
Точность = 99,89%

ROC AUC = 99,25%

F1-score для класса 0 = 100%

F1-score для класса 1 = 98%

Модель для полного изображения



Спасибо за внимание!