

Лабораторная работа

Методы классификации

Работу подготовили:

Панов Олег, Михаил Бабушкин, Денис Чашин, Анатолий Мезенов, Никита Бабушкин

Входные данные

- Дана база из ~1500 сайтов.
- Сайт может быть фишинговым, не фишинговым или подозрительным [1, -1, 0]
- У Каждого сайта есть 9 атрибутов.

Задача

- Классифицировать сайты по нашим параметрам на фишинговые, подозрительные и не фишинговые

Натренировать модель классификации данных, выбрать лучшую из тренированных и показать результат

Наш подход

Для решения проблемы классификации мы решили выбрать следующие модели, и распределили их между собой.

- LDA - Михаил Бабушкин
- GNB - Денис Чашин
- DTC - Никита Боровик
- KNN - Олег Панов
- SVC - Анатолий Мезенов

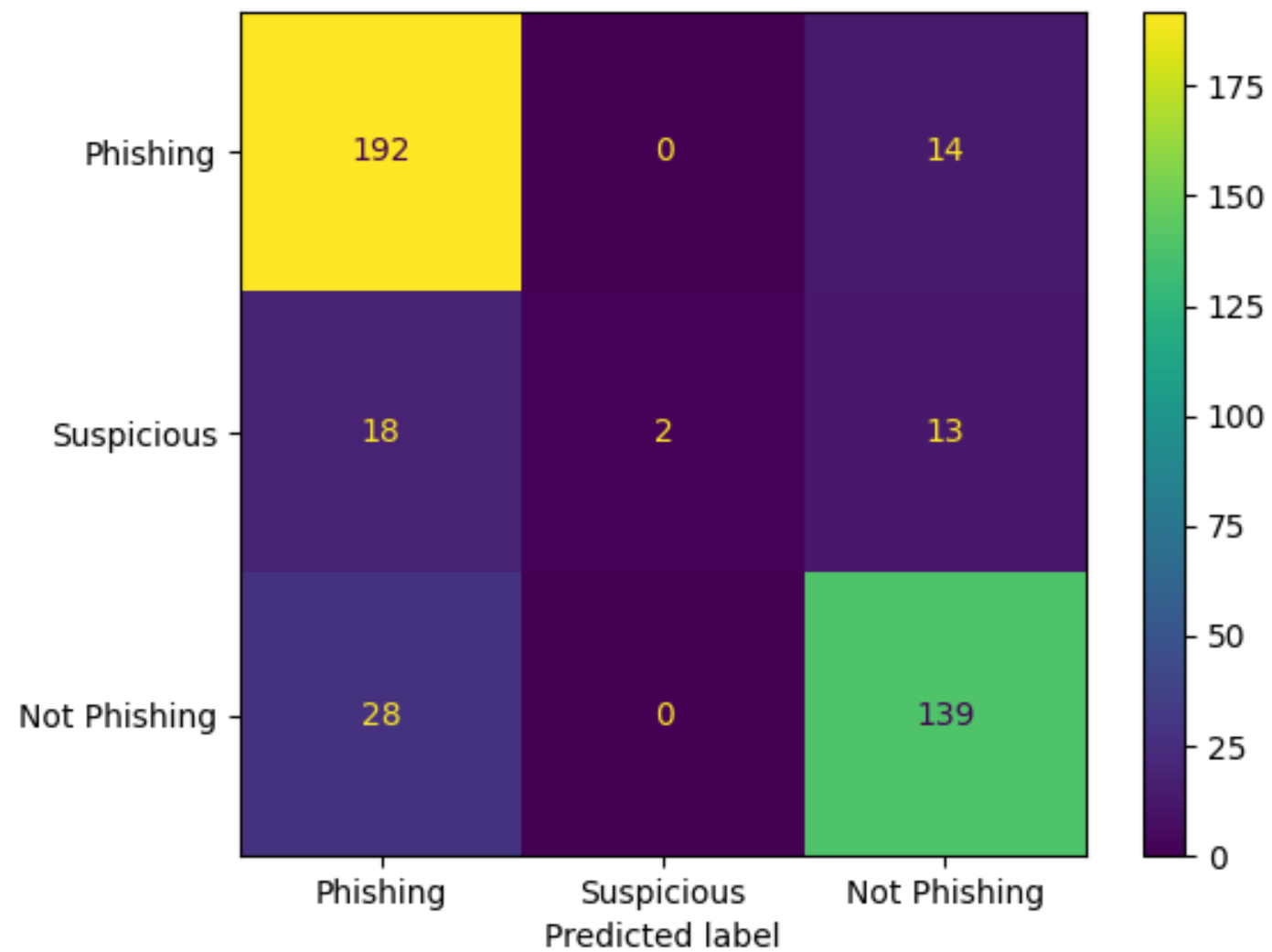
Модель LDA

Linear Discriminant Analysis (линейный дисериминантный анализ)

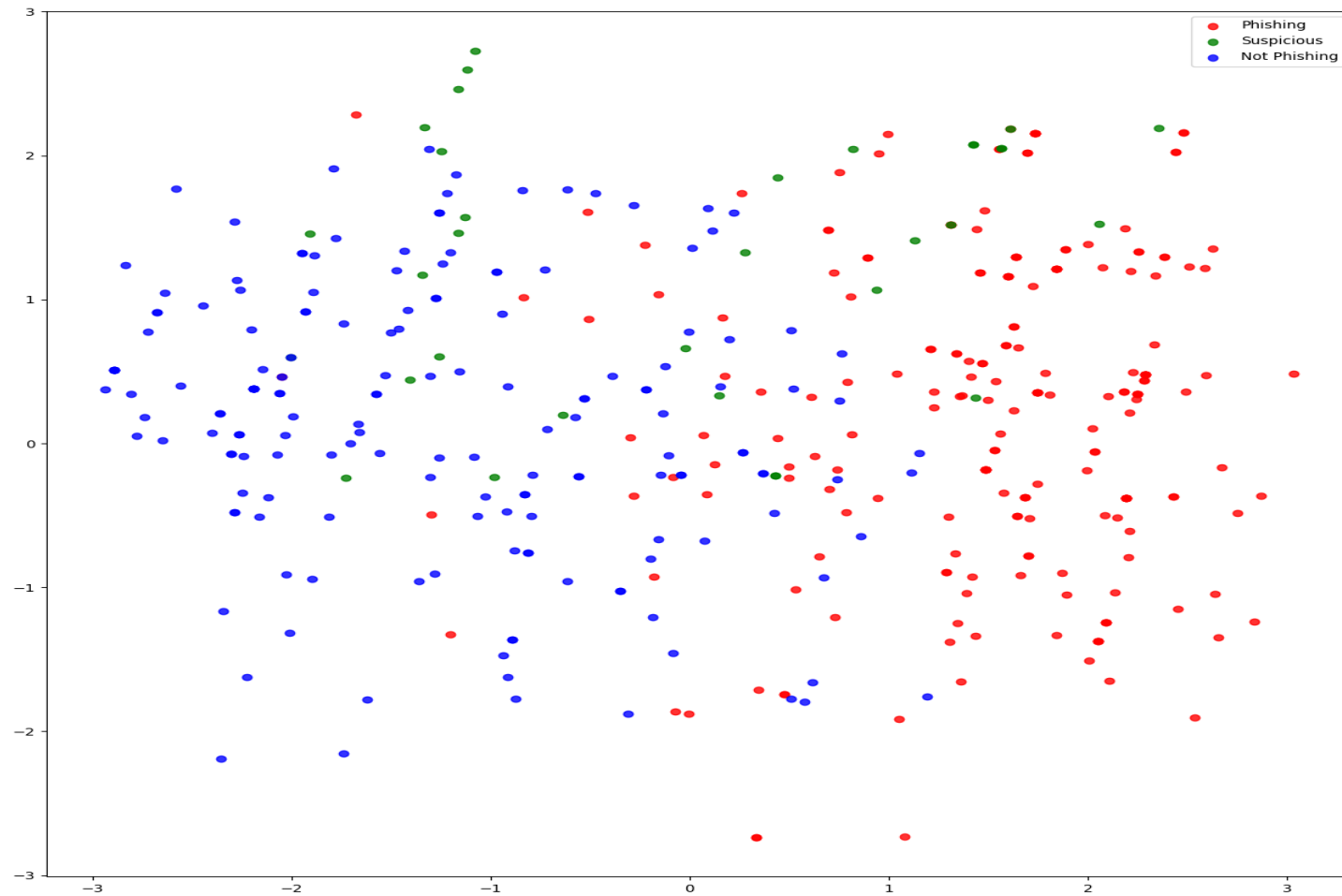
Модель LDA представляет собой ...

Она хороша потому что...

Но также...



Matrix for LDA model



LDA visualisation

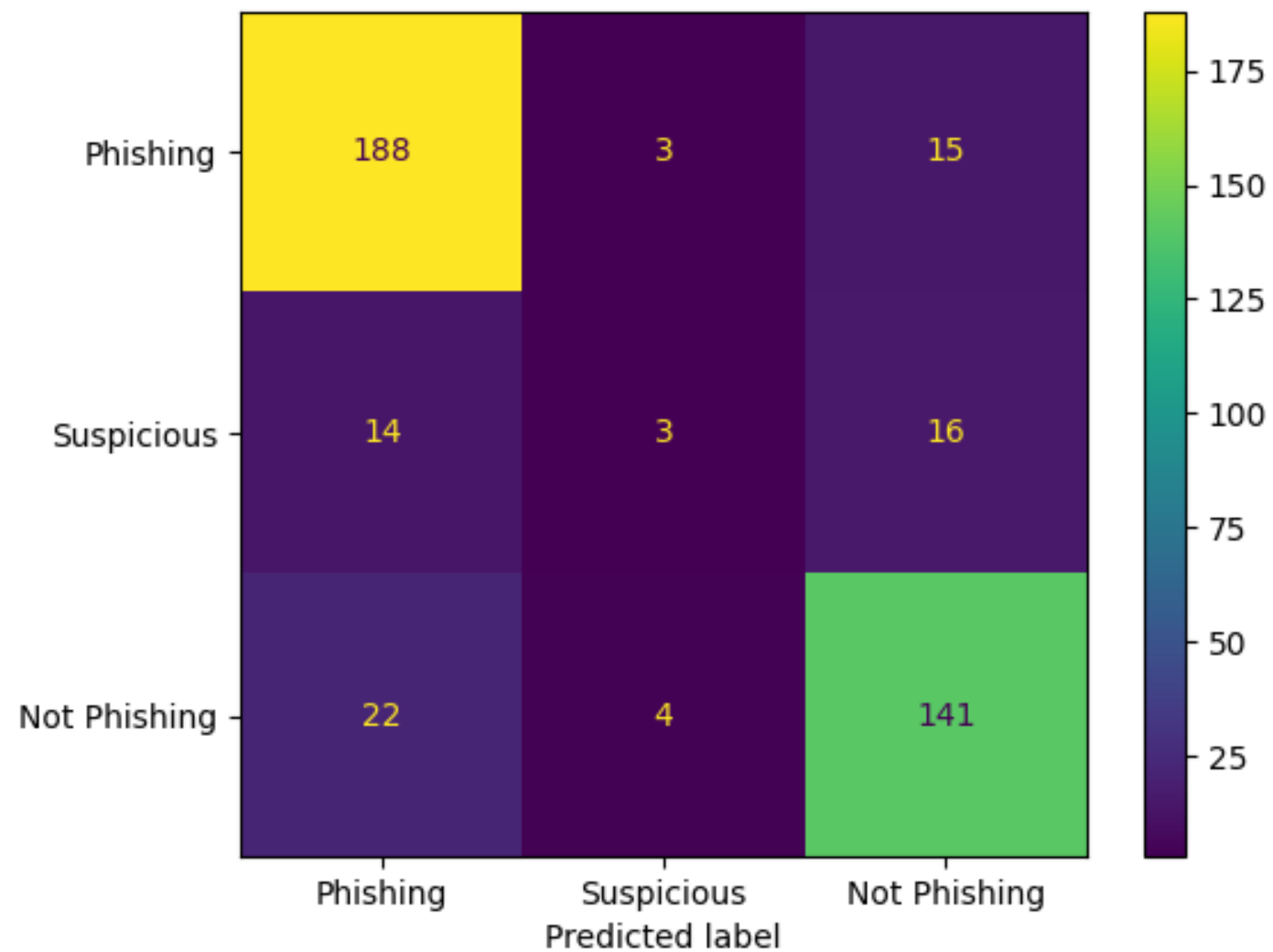
Модель GNB

| Gaussian Naive Baías (наивный байесовский классификатор)

Модель GNB представляет собой...

Она удобна в...

Но...



Matrix for GNB model

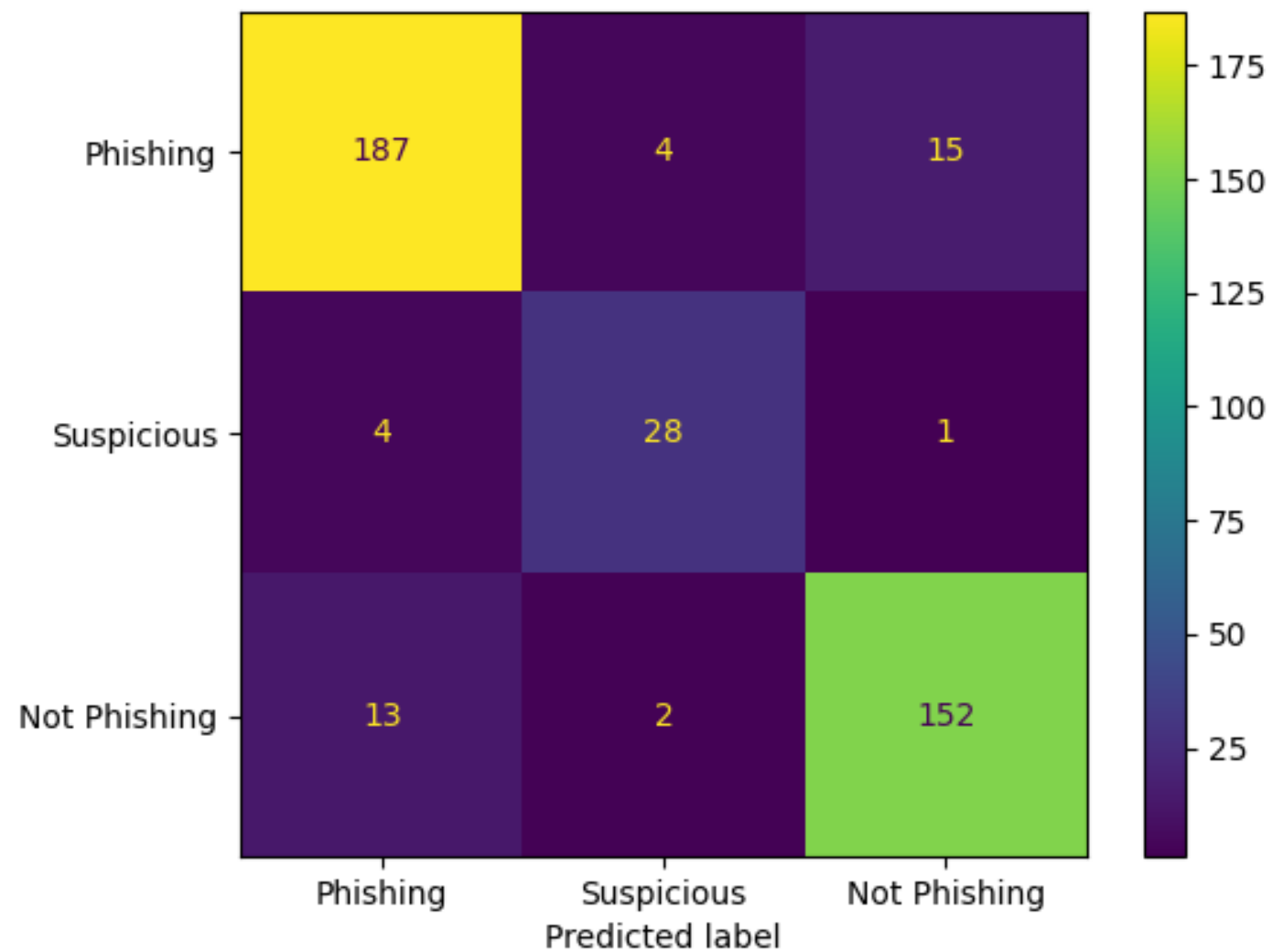
Моедль DTC

Decision Tree Classifier (Классификатор дерева решений)

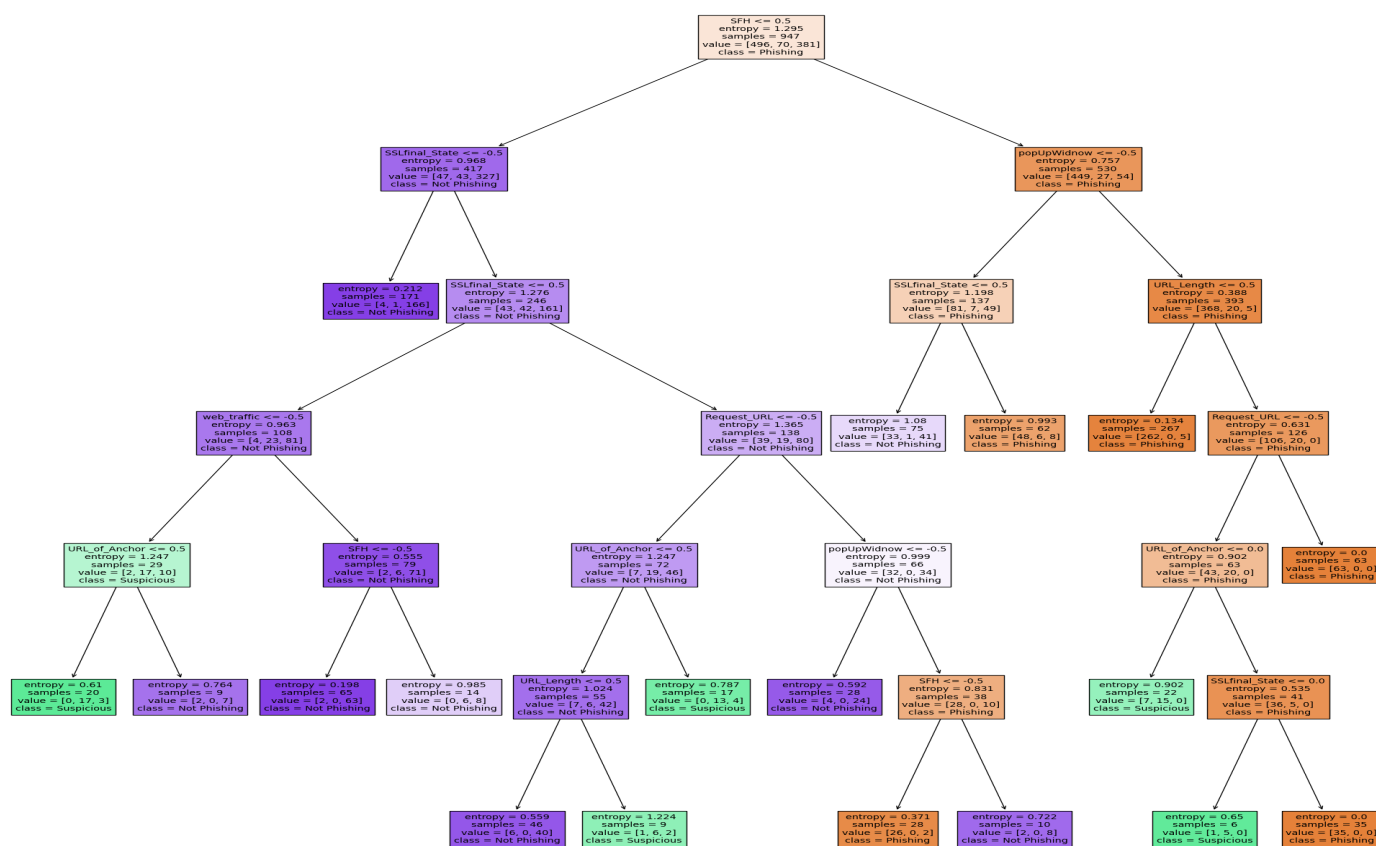
Модель DTC представляет собой... _(кратко)

Она имеет...

Но еще...



Matrix for DTC model

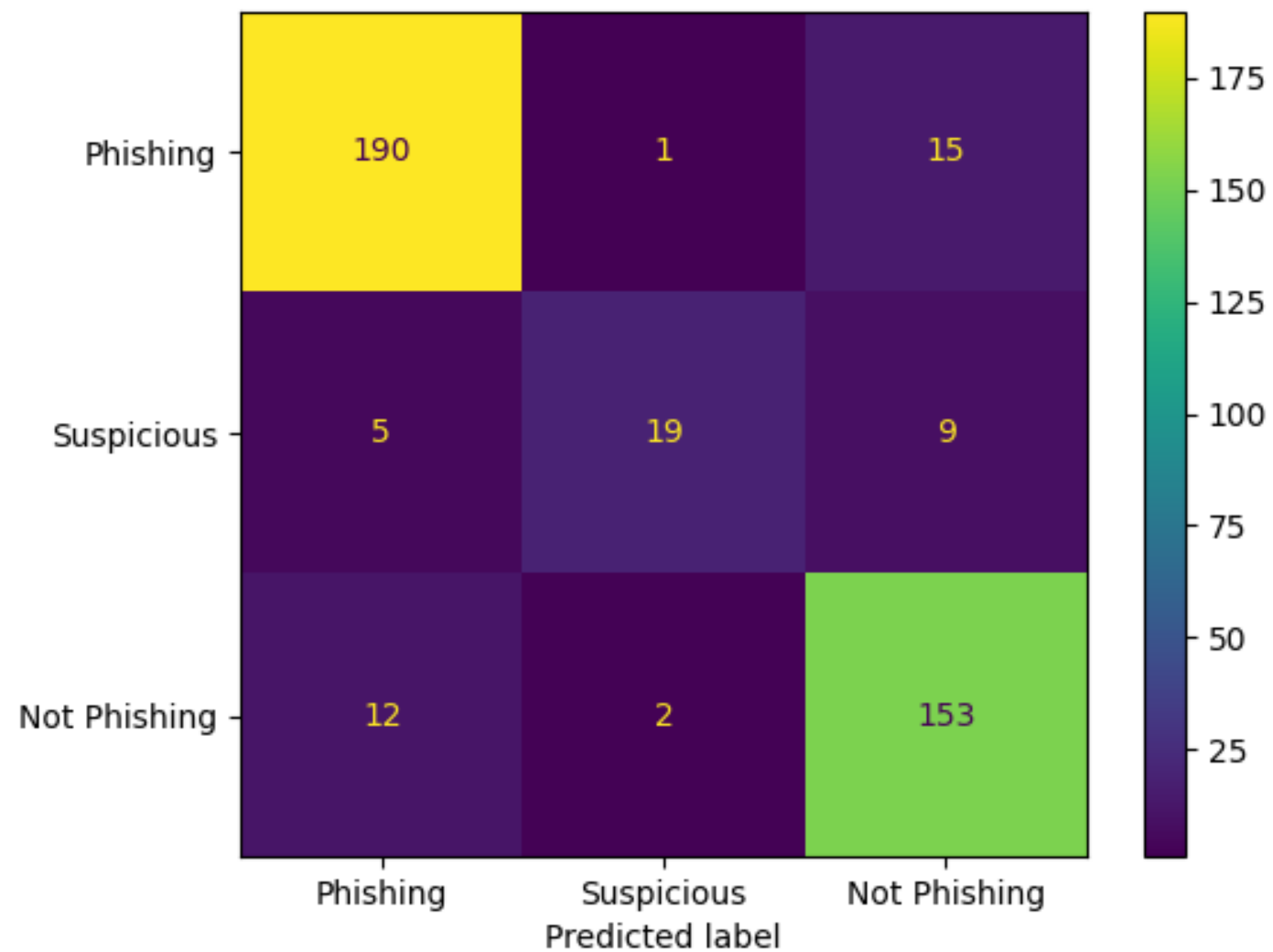


Desision Tree visualisation

Модель KNN

K Nearest Neighbors (классификатор "k-ближайших соседей")

- Модель KNN представляет собой распределение элементов (*сайтов*) в плоскости и просчет расстояния до k его ближайших элементов (соседей). И именно по преимуществу соседей принадлежащих множеству определяется тип текущего элемента
- Модель хороша, тк ее метод не использует сложную математику и реализация сразу понятна



Matrix for KNN model

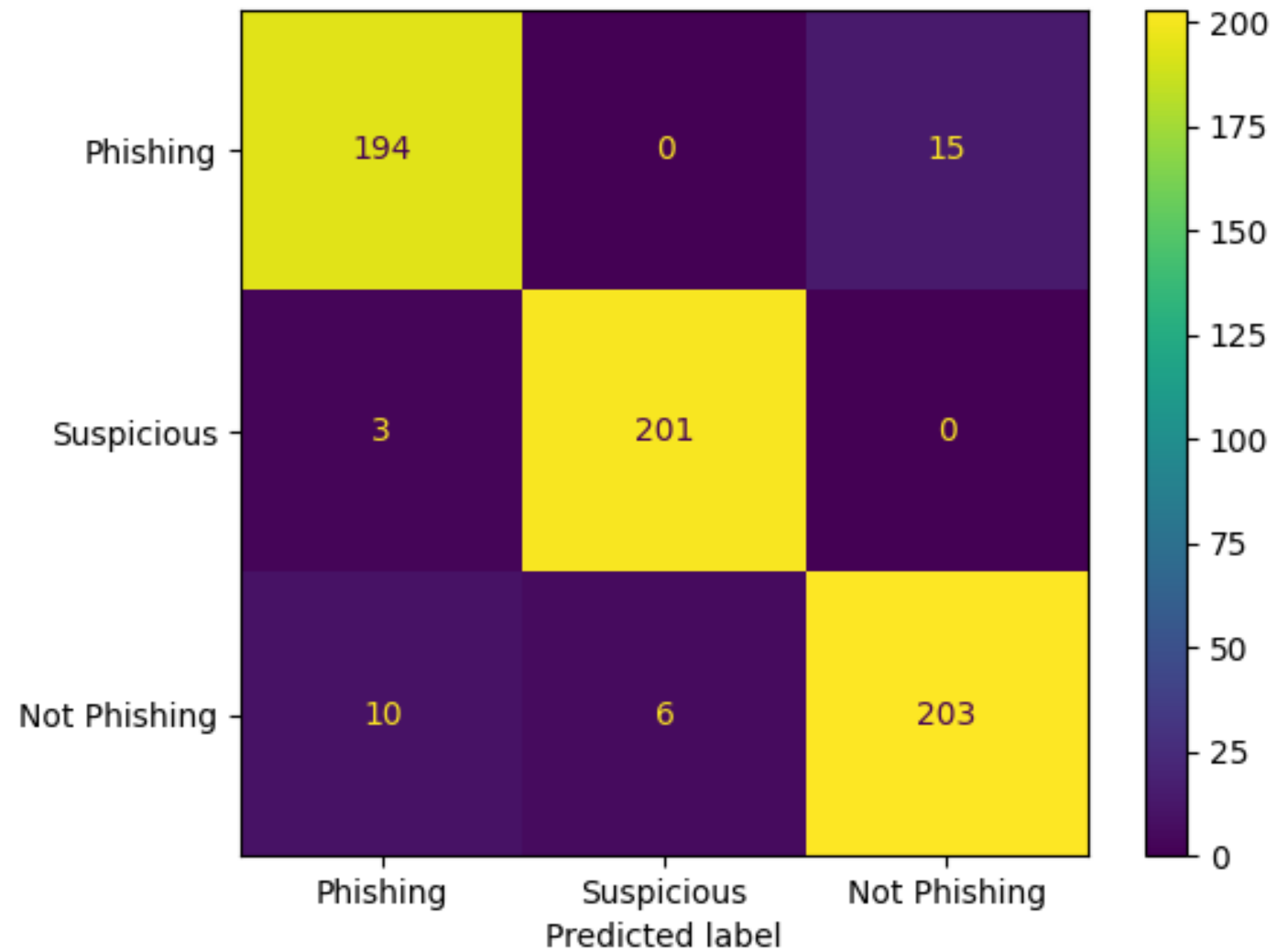
Моедль SVC

| Support Vector Classification (метод опорных векторов)

Модель SVC представляет собой... _(кратко)

Она имеет...

Но еще...



Matrix for SVC model

Итоги

лол

Спасибо за внимание