Разработка модели по определению половой принадлежности на основании аудиозаписи голоса

Выполнила: Автухович Надежда Сергеевна
Руководитель: Инженер-исследователь, Сельский Виктор

Актуальность

Актуальность обусловлена стремительным ростом востребованности бесконтактных и пассивных методов биометрической идентификации. В отличии от традиционных способов верификации, анализ голоса позволяет определить половую принадлежность пользователя без его активного участия и дополнительных действий, что важно для систем непрерывной аутентификации, голосовых помощников и интеллектуальных сервисов.

Цель и задачи исследования

Цель

Разработать модель машинного обучения для определения половой принадлежности человека на основании аудиозаписи голоса.

Задачи

1. Провести анализ предметной области.
2. Сформировать и подготовить датасет.
3. Реализовать и обучить модель машинного обучения.
4. Провести тестирование и оценить точность работы.

Акустическая структура голоса

Анатомическая причина

Мужской голос

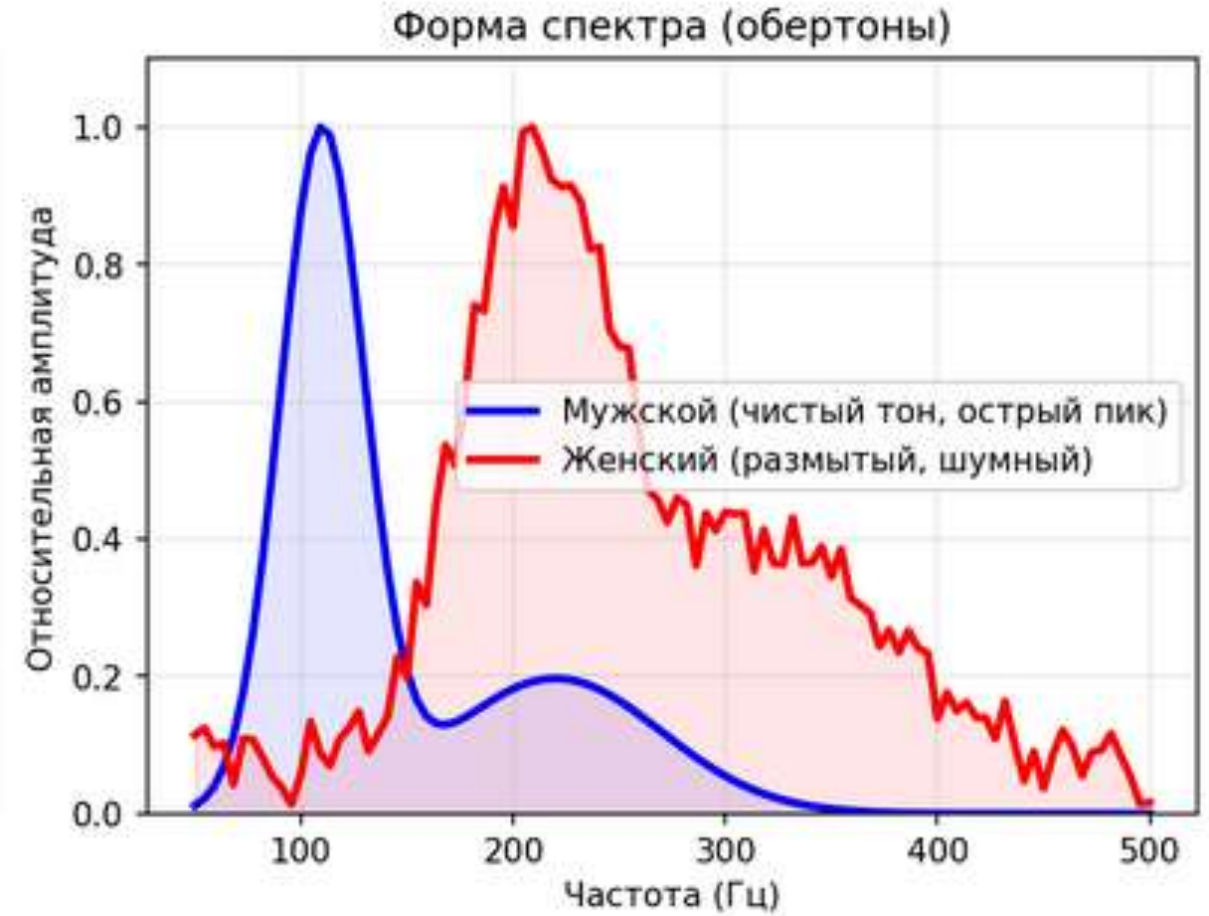
Связки: длинные и толстые

→ F0: 85 – 180 Гц

Женский голос

Связки: короткие и тонкие

→ F0: 165 – 255 Гц



Ключевые акустические маркеры пола

Характеристики	**Мужской голос**	**Женский голос**
Фундаментальная частота (F0)	Ниже (80 - 180 Гц)	Выше (165 - 255 Гц)
Средняя частота спектра	Ниже	Выше
Спектральный центр тяжести	Смещен в низкочастотную область	Смещен в высокочастотную область
Спектральная энтропия	Ниже	Выше
Тембр	Более низкий	Более высокий
Изменчивость интонации	Ниже	Выше

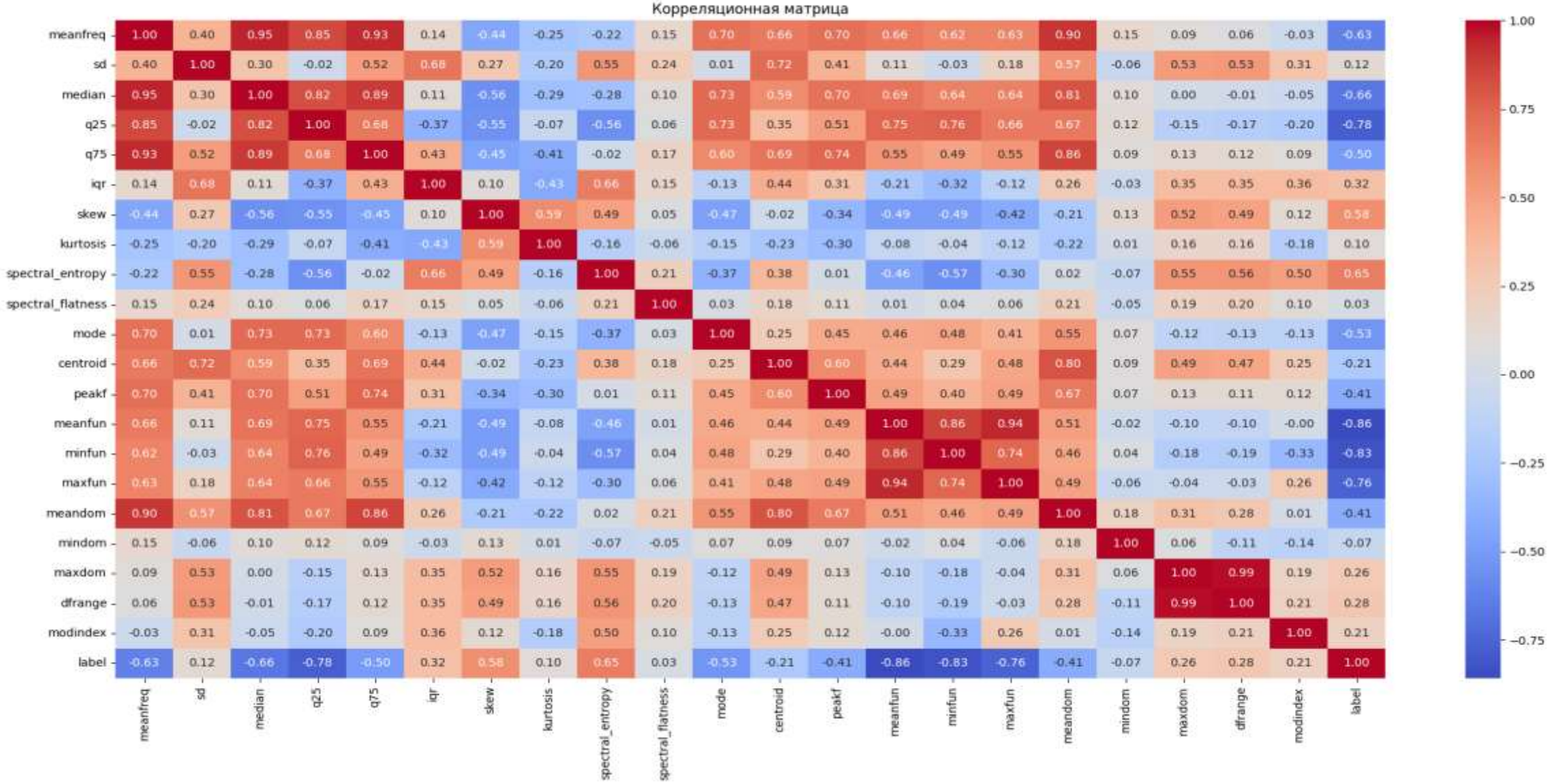
Данные для анализа

	meanfreq	sd	median	q25	q75	iqr	skew	kurtosis	spectral_entropy	spectral_flatness	mode	centroid	peakf	meanfun	minfun
0	0.2166	0.0614	0.2109	0.1953	0.2578	0.0625	-0.3237	1.4458	0.7808	0.4826	0.2002	0.2366	0.2031	0.2223	0.1757
1	0.1346	0.0298	0.1406	0.1250	0.1484	0.0234	-0.1346	0.9467	0.7656	0.4097	0.1474	0.1889	0.1484	0.1446	0.1082

maxfun	meandom	mindom	maxdom	dfrange	modindex	label
0.3167	0.2309	0.0625	0.3984	0.3359	0.1388	female
0.1788	0.1519	0.0703	0.3906	0.3203	0.1010	male

1. Основные частотные характеристики (meanfreq, sd, median, q25, q75, iqr, skew, kurtosis, mode, centroid, peakf)
2. Параметры «Фундаментальной частоты» (meanfun, minfun, maxfun)
3. Параметры «Доминантной частоты» (meandom, mindom, maxdom, dfrange)
4. Спектральные характеристики (spectral_entropy, spectral_flatness, modindex)

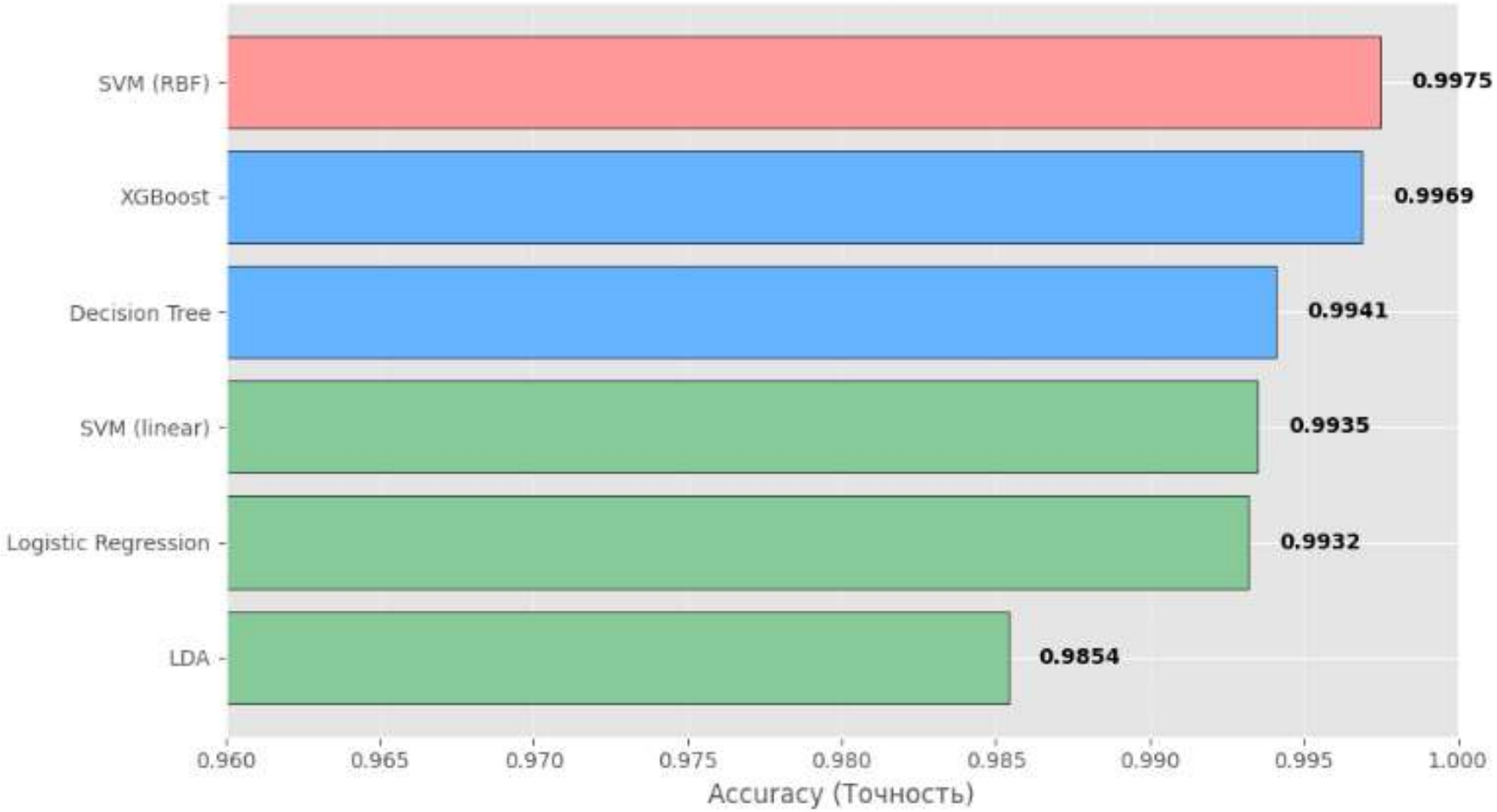
Корреляционная матрица



Сравнительный анализ моделей классификации

Модель	Accuracy	Precision	Recall	MSE
Линейные модели				
SVM(linear)	0.9935	0.9957	0.9942	0.0065
Logistic Regression	0.9854	0.9957	0.9937	0.0068
LDA	0.9854	0.9806	0.9971	0.0146
Нелинейные модели				
SVM (RBF)	0.9975	0.9976	0.9968	0.0025
XGBoost	0.9969	0.9986	0.9966	0.0031
Decision Tree	0.9941	0.9971	0.9937	0.0059

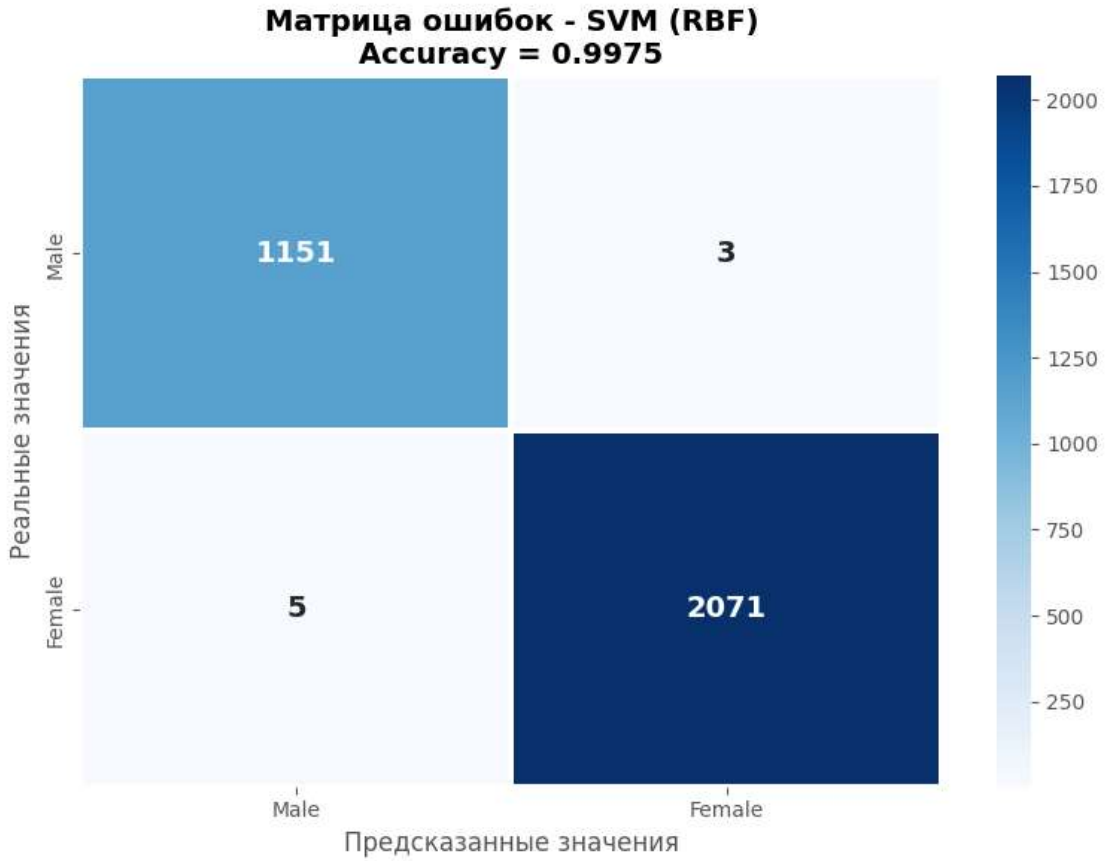
Сравнение точности работы моделей



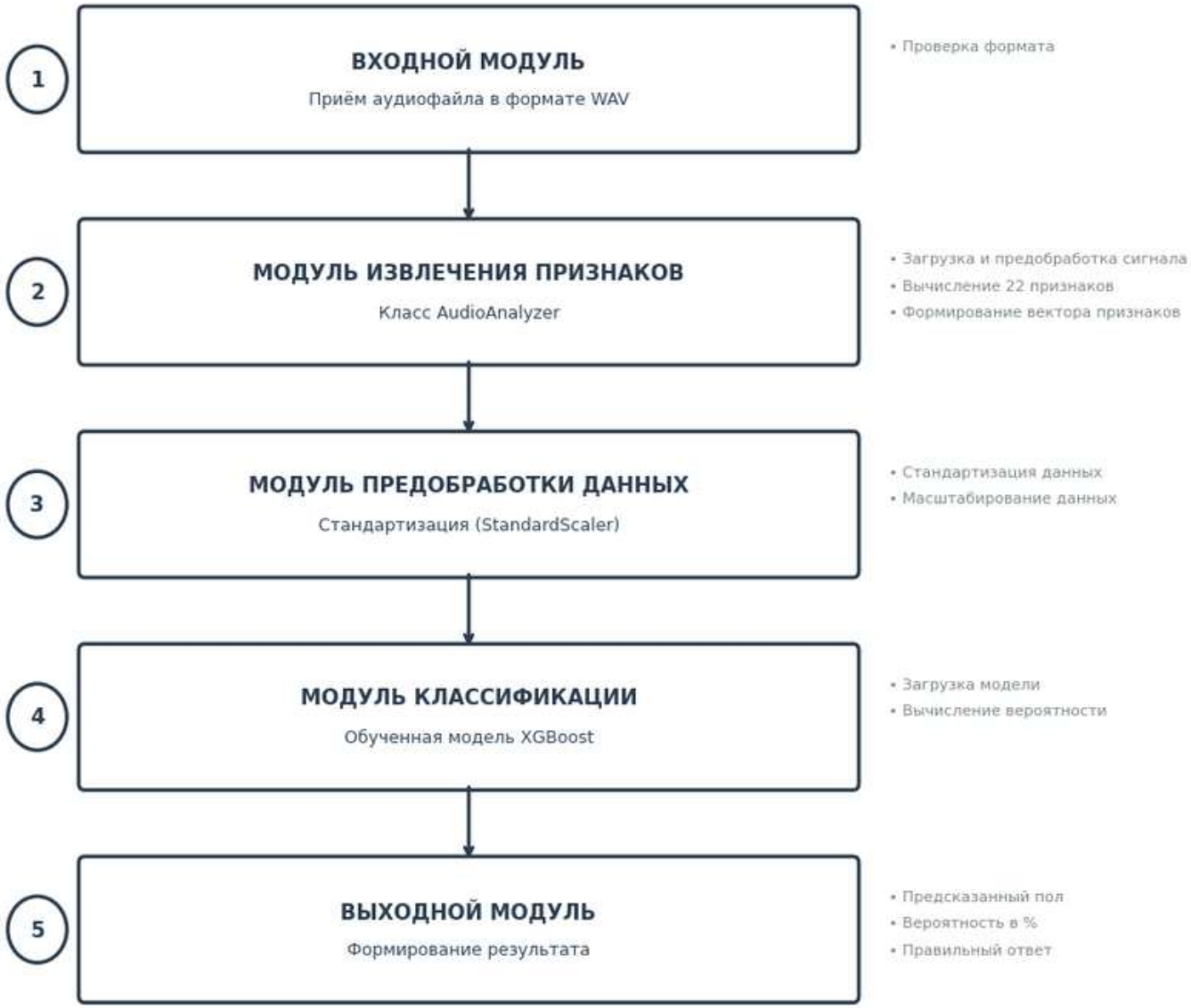
Лучше всего себя показала модель SVM (RBF)

Прогнозирование на основе лучшей модели

	precision	recall	f1-score	support
Male	0.9957	0.9974	0.9965	1154
Female	0.9986	0.9976	0.9981	2076
accuracy			0.9975	3230
macro avg	0.9971	0.9975	0.9973	3230
weighted avg	0.9975	0.9975	0.9975	3230

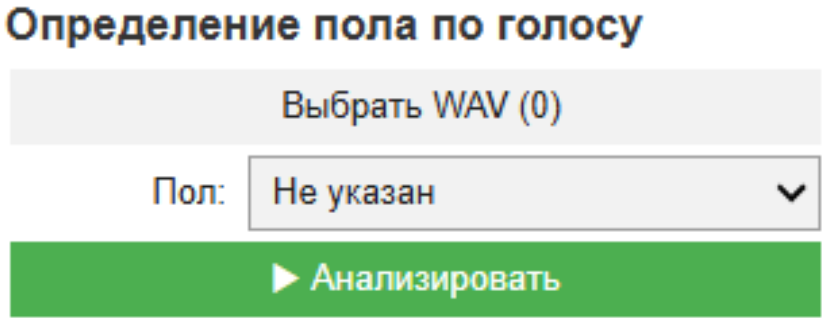


Архитектура модели определения пола по голосу

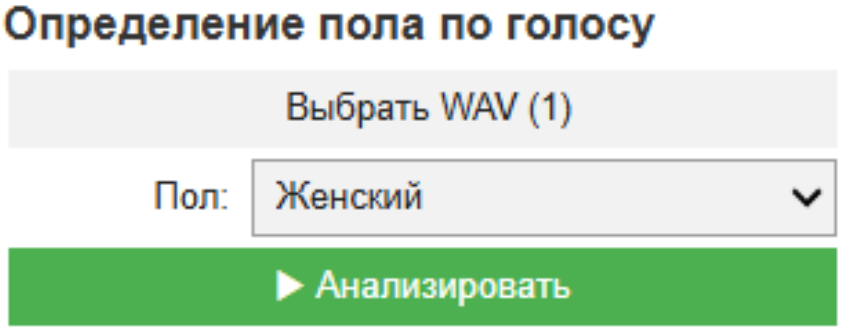


Тестирование

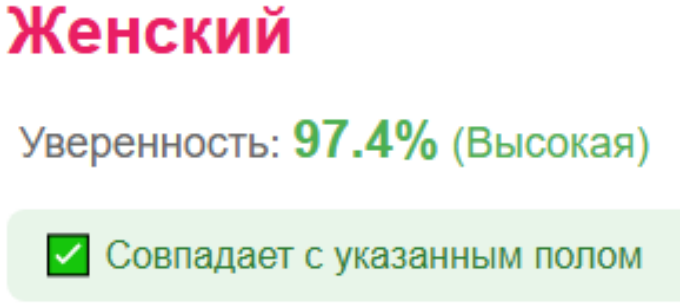
1. Загрузить аудиофайл и указать пол говорящего



2. Нажать на кнопу Анализировать



3. В результате будет выведен:
   - Предсказанный пол;
   - Уверенность модели в правильности предсказания;
   - Уведомление о том, совпадает ли ожидаемый результат с предсказанным или нет.



Заключение

- Была синтезирована модель по определению половой принадлежности на основании аудиозаписи голоса;
- Оптимальной моделью была выбрана SVM с RBF – ядром из-за наилучших результатов;
- Было проведено тестирование работы разработанной модели.