Developing informational system for collecting and analyzing student’s digital footprint

Author: Moskalev D.I.

NSTU, 2022

ЦЕЛИ И ЗАДАЧИ РАБОТЫ

Цели:

- Извлечение данных из группового журнала в сфотографированном или отсканированном виде;
- Обработка и запись в базу данных BigQuery (ЦИУ);
- Визуализация данных в виде панели визуализации (дашборда) и веб-сервиса.

Задачи:

- Построение модели классификации рукописных символов;
- Создание мобильных приложений для сбора, обработки и записи данных;
- Визуализация результатов.

ЗАЧЕМ НУЖЕН СБОР И АНАЛИЗ ДАННЫХ?

Цифровой след студента – полученные обезличенные цифровые данные студента из группового журнала, состоящие из уникального номера и количества посещений (пропусков) учебных занятий за определенный период.

Сбор и анализ данных цифрового следа студента позволяет:

- Анализировать текущее состояние студента, группы, факультета или университета в целом;
- Персонализировать рекомендации для отдельных сегментов учащихся в университете.

ПОСТАНОВКА ЗАДАЧИ

Формальная постановка задачи классификации представляет собой неизвестную целевую зависимость

$$y^*: X \to Y, \qquad (1)$$

где X – множество описаний входных объектов, Y – множество классов.

При этом значения отображения (1) известны только на объектах обучающей выборки:

$$X_n = [(x_1, y_1), \dots, (x_n, y_n)],$$

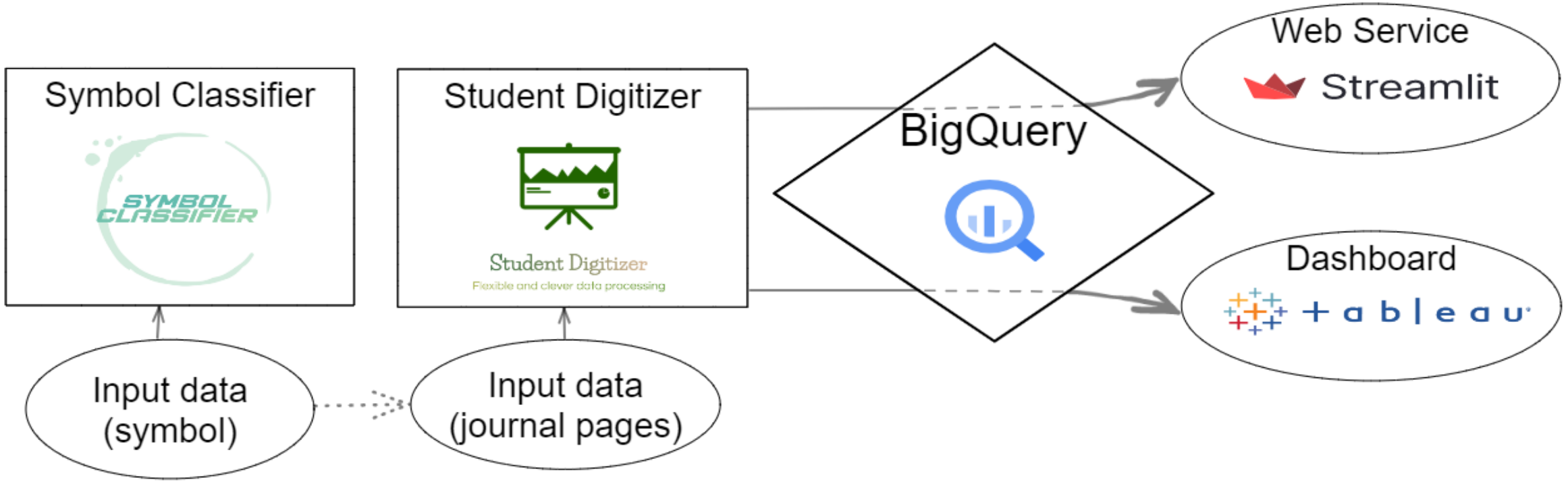
где n – количество строк объектов.

В задаче многоклассовой классификации (мульти-классификации) множество классов:

$$Y = \{f_1, f_2, \dots, f_n\}, \text{где } f_1 = 0, f_2 = 1, \dots, f_n = n - 1.$$

ВЫБРАННЫЕ МЕТОДЫ РЕШЕНИЯ

Для построения модели данных были выбраны многослойные сверточные нейронные сети. В качестве основного метода решения задачи была сделана ETL система по сбору, обработке и загрузке данных.



МОБИЛЬНОЕ ПРИЛОЖЕНИЕ “SYMBOL CLASSIFIER”

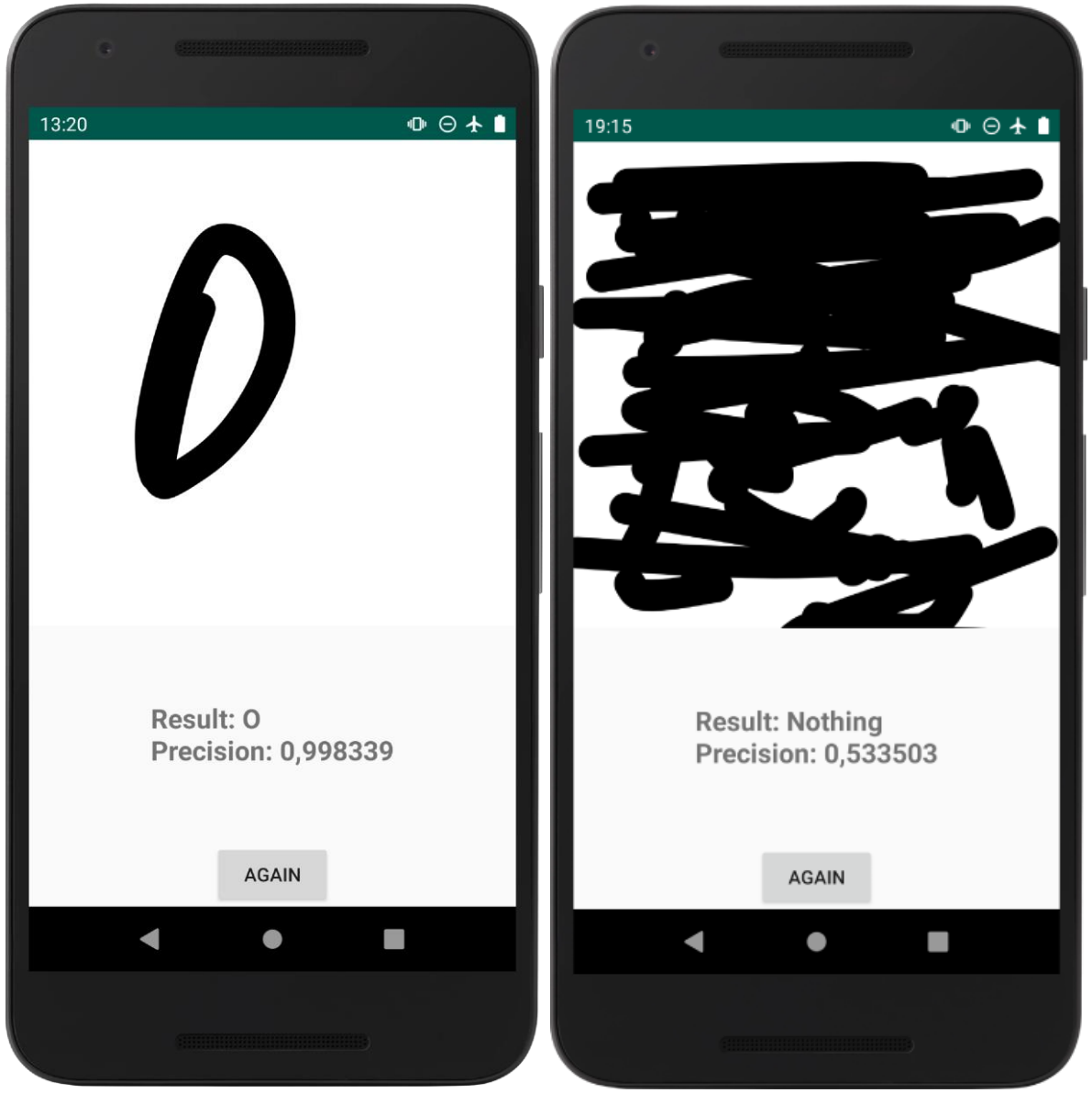
Мобильное приложение для классификации рукописных символов.

В приложении 4 класса: Б, Н, О и “Nothing”.

Пример для классов: Б, Н.

Условие принадлежности к классу “Nothing”: символ для всех классов букв имеет метрику точности менее 0.75.

МОБИЛЬНОЕ ПРИЛОЖЕНИЕ “SYMBOL CLASSIFIER”



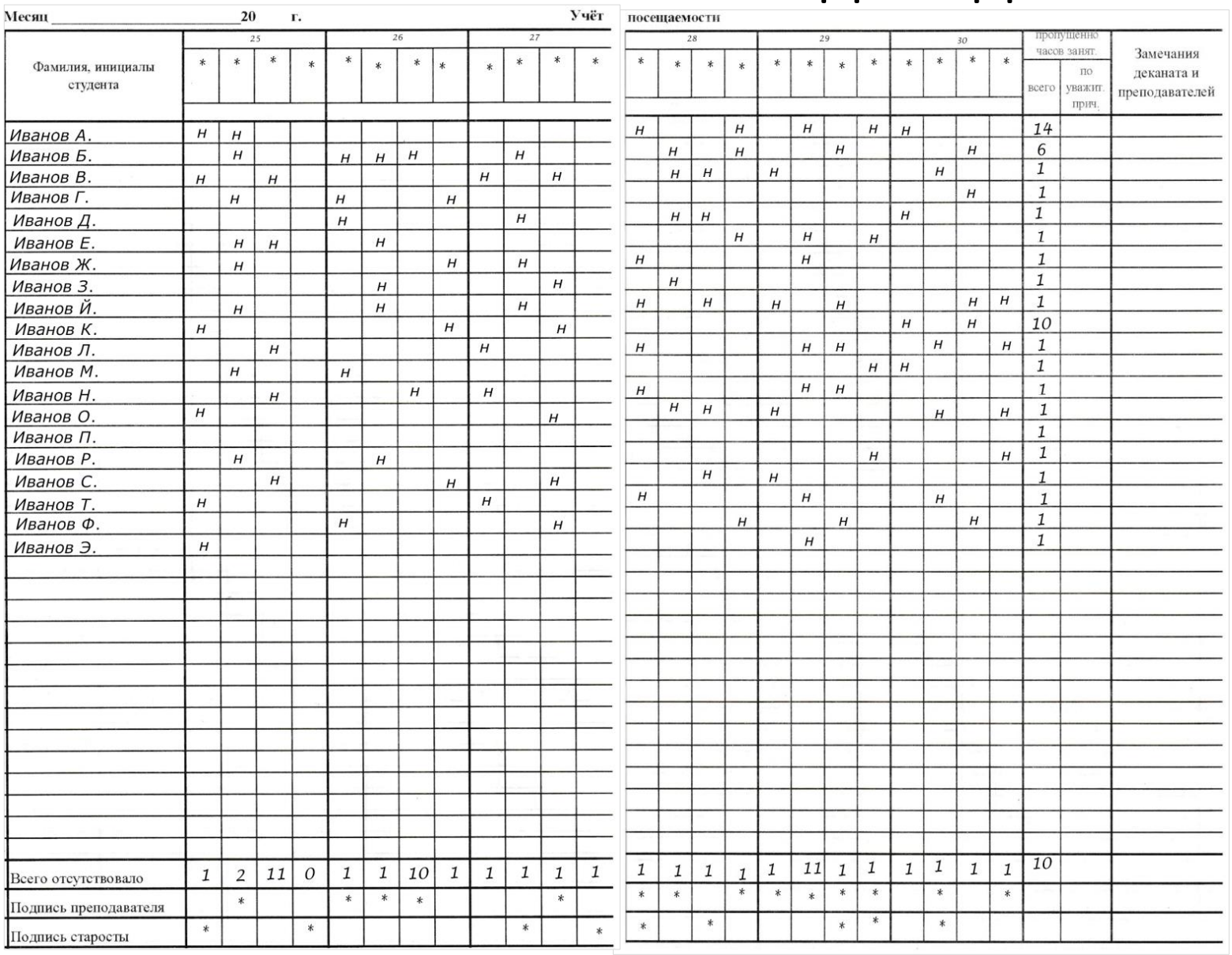
Мобильное приложение для классификации рукописных символов.

В приложении 4 класса: Б, Н, О и “Nothing”.

Пример для классов: О, Nothing.

Условие принадлежности к классу “Nothing”: символ для всех классов букв имеет метрику точности менее 0.75.

ВХОДНЫЕ ДАННЫЕ



Месяц _____ 20 г. Учёт посещаемости

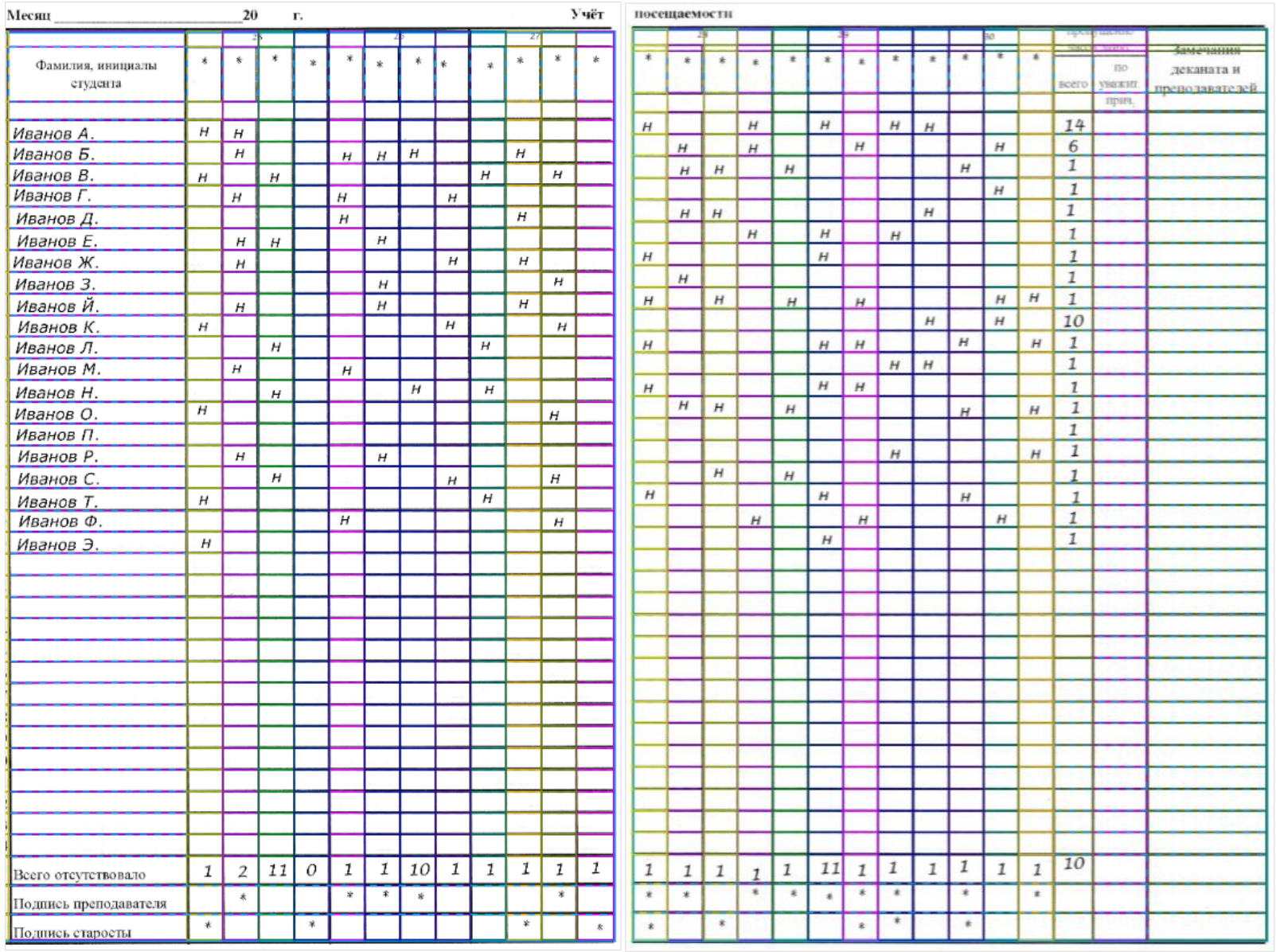
Фамилия, инициалы студента	25				26				27				28				29				30				пропущено часов занят. всего	пропущено часов занят. по уважит. прич.	Замечания деканата и преподавателей
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*			
Иванов А.	Н	Н											Н			Н		Н		Н	Н				14		
Иванов Б.		Н			Н	Н	Н			Н				Н		Н			Н				Н		6		
Иванов В.	Н		Н						Н		Н			Н	Н		Н					Н			1		
Иванов Г.		Н			Н			Н															Н		1		
Иванов Д.					Н					Н				Н	Н						Н				1		
Иванов Е.		Н	Н			Н										Н		Н		Н					1		
Иванов Ж.		Н						Н		Н			Н					Н							1		
Иванов З.						Н					Н			Н											1		
Иванов Й.		Н				Н				Н			Н		Н		Н				Н	Н			1		
Иванов К.	Н							Н			Н										Н		Н		10		
Иванов Л.			Н						Н				Н				Н	Н				Н		Н	1		
Иванов М.		Н			Н														Н	Н					1		
Иванов Н.			Н				Н		Н				Н				Н	Н							1		
Иванов О.	Н													Н	Н		Н					Н		Н	1		
Иванов П.																									1		
Иванов Р.		Н				Н													Н					Н	1		
Иванов С.			Н					Н						Н			Н								1		
Иванов Т.	Н								Н				Н					Н				Н			1		
Иванов Ф.					Н										Н				Н						1		
Иванов Э.	Н																Н								1		
Всего отсутствовало	1	2	11	0	1	1	10	1	1	1	1	1	1	1	1	1	1	11	1	1	1	1	1	1	10		
Подпись преподавателя		*			*	*	*				*		*	*		*	*	*	*	*		*		*			
Подпись старосты	*			*						*		*	*		*				*	*		*					

Входные данные представлены в виде 2 страниц группового журнала (отсканированная версия).

На этапе обработки данных страницы соединяются в 1 лист и производится обработка данных на предмет соответствия типов данных, предметов из расписания и подсчета пропусков предмета студентами.

Данные журнала вымышленные.

РЕЗУЛЬТАТЫ ОПТИЧЕСКОГО РАСПОЗНАВАНИЯ ТЕКСТА



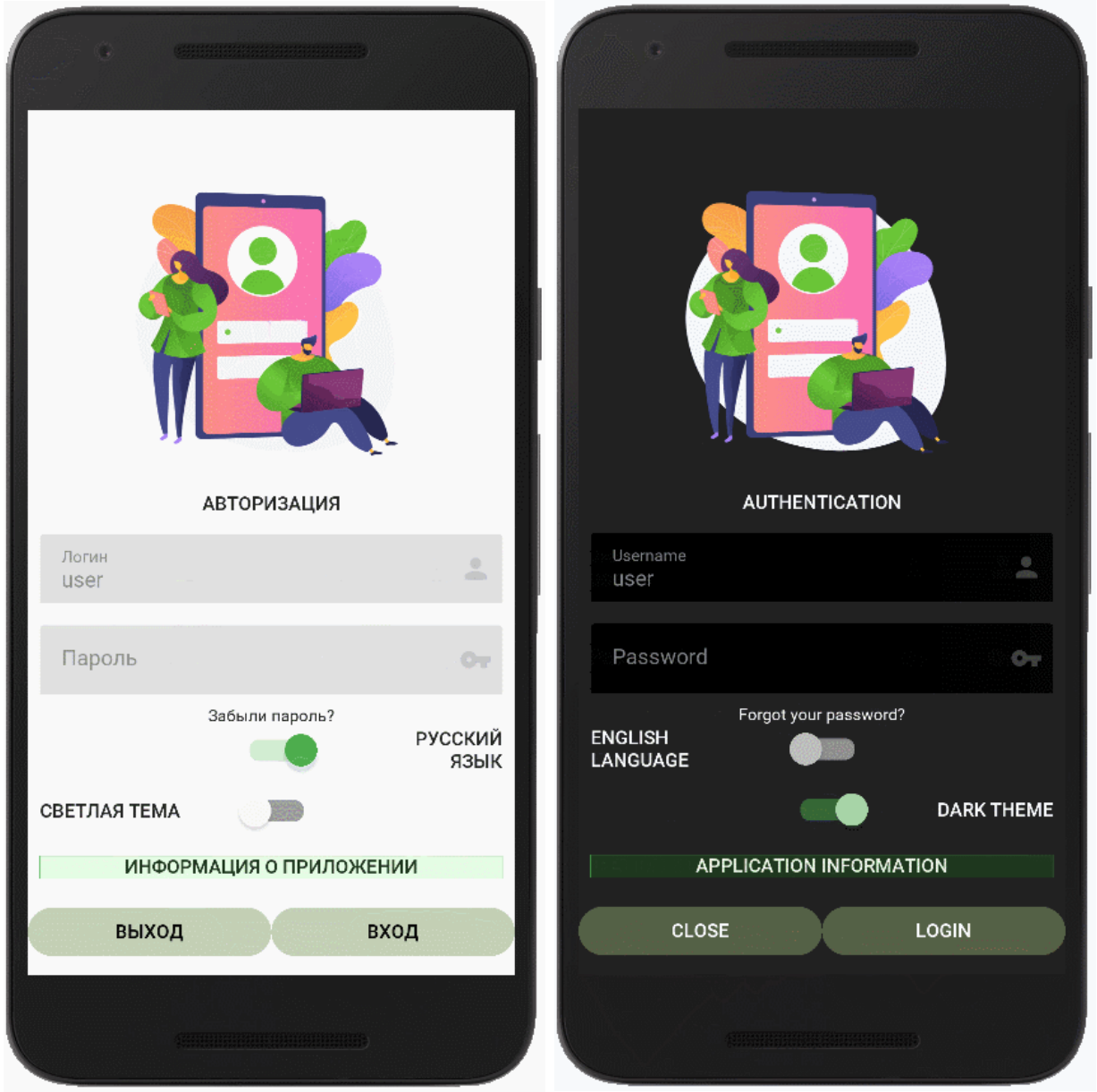
Месяц _____ 20 г. Учёт посещаемости

Фамилия, инициалы студента	25				26				27				28				всего	по уважит. прич.	Замечания деканата и преподавателей
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*			
Иванов А.	Н	Н											Н			Н	14		
Иванов Б.		Н			Н	Н	Н			Н				Н		Н	6		
Иванов В.	Н		Н						Н		Н		Н	Н			1		
Иванов Г.		Н			Н			Н							Н		1		
Иванов Д.					Н					Н			Н				1		
Иванов Е.		Н	Н			Н								Н			1		
Иванов Ж.		Н						Н		Н							1		
Иванов З.						Н					Н						1		
Иванов Й.		Н				Н				Н							1		
Иванов К.	Н							Н			Н			Н	Н		10		
Иванов Л.			Н						Н				Н		Н		1		
Иванов М.		Н			Н								Н	Н			1		
Иванов Н.			Н				Н		Н								1		
Иванов О.	Н										Н				Н		1		
Иванов П.																	1		
Иванов Р.		Н				Н									Н		1		
Иванов С.			Н					Н			Н						1		
Иванов Т.	Н								Н					Н			1		
Иванов Ф.					Н						Н				Н		1		
Иванов Э.	Н																1		
Всего отсутствовало	1	2	11	0	1	1	10	1	1	1	1	1	1	1	1	1	10		
Подпись преподавателя		*			*	*	*				*		*		*				
Подпись старосты	*			*						*		*	*			*			

Результаты оптического распознавания текста в таблицах группового журнала используя сервис оптического распознавания символов (OCR) Nanonets.

Данные журнала вымышленные.

МОБИЛЬНОЕ ПРИЛОЖЕНИЕ “STUDENT DIGITIZER”



Основное мобильное приложение для сбора, обработки и перемещения данных.

В мультиязычном приложении реализованы 2 темы.

На слайде представлен вариант с авторизацией пользователя и проверкой данных модели из сервиса распознавания символов, обработкой данных с промежуточным визуальным представлением и экспортом полученных таблиц в базу данных BigQuery.

GIF файлы:
https://github.com/EnterSub/Student-Digitizer

ВЫХОДНЫЕ ДАННЫЕ

Выходные данные для записи в базу данных представлены в виде 2 таблиц группового журнала.

1 таблица: таблица данных по студентам для записи в БД BigQuery.

	number	student	lectures_all	group	week_n	date
0	1	Student1	2	ПММ-03	7	2022-06-18
1	2	Student2	5	ПММ-03	7	2022-06-18
2	3	Student3	2	ПММ-03	7	2022-06-18
3	4	Student4	2	ПММ-03	7	2022-06-18
4	5	Student5	2	ПММ-03	7	2022-06-18
5	6	Student6	2	ПММ-03	7	2022-06-18
6	7	Student7	2	ПММ-03	7	2022-06-18
7	8	Student8	1	ПММ-03	7	2022-06-18
8	9	Student9	3	ПММ-03	7	2022-06-18
9	10	Student10	1	ПММ-03	7	2022-06-18
10	11	Student11	1	ПММ-03	7	2022-06-18
11	12	Student12	2	ПММ-03	7	2022-06-18
12	13	Student13	2	ПММ-03	7	2022-06-18
13	14	Student14	1	ПММ-03	7	2022-06-18
14	15	Student15	0	ПММ-03	7	2022-06-18
15	16	Student16	2	ПММ-03	7	2022-06-18
16	17	Student17	0	ПММ-03	7	2022-06-18
17	18	Student18	2	ПММ-03	7	2022-06-18
18	19	Student19	1	ПММ-03	7	2022-06-18
19	20	Student20	1	ПММ-03	7	2022-06-18

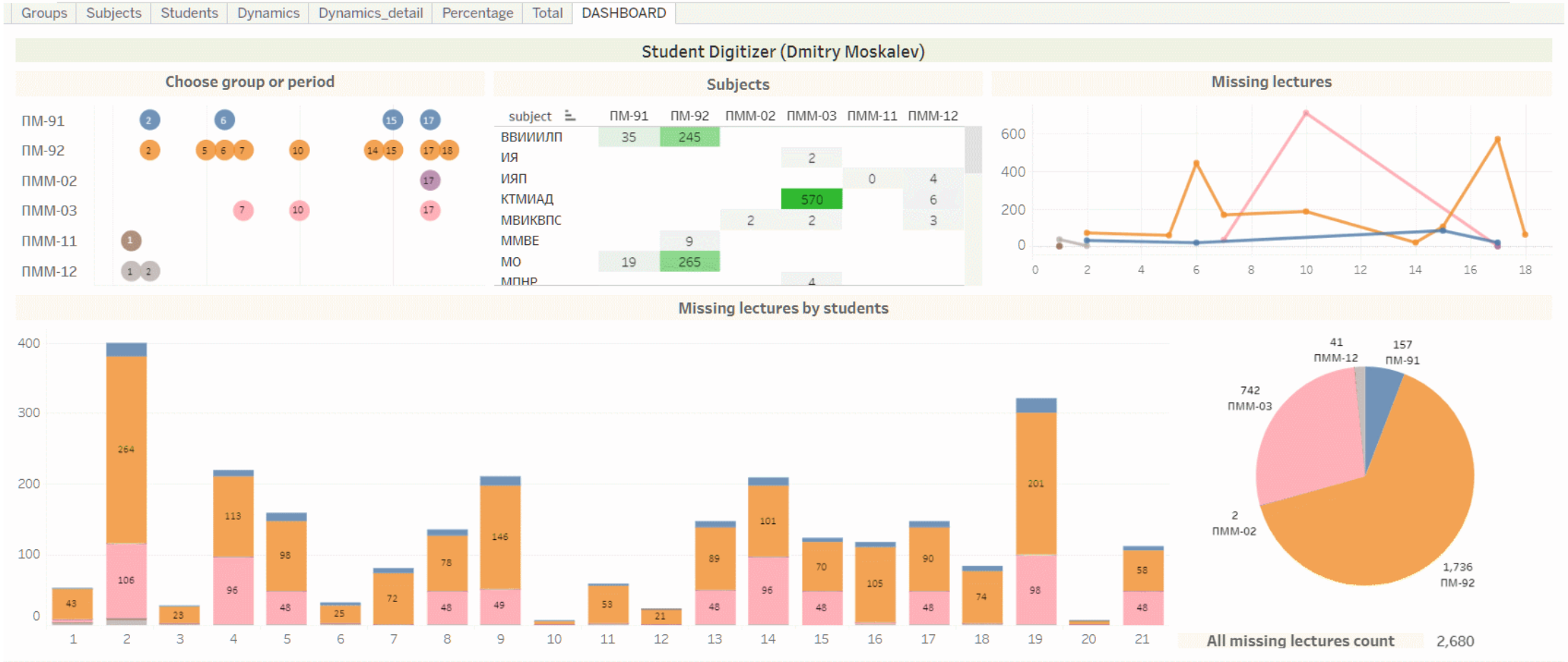
2 таблица: таблица данных по предметам (аббревиатурам) для записи в БД BigQuery.

	subject	group	date	week_n	total
0	ИЯ	ПММ-03	2022-06-18	7	6
1	КТМИАД	ПММ-03	2022-06-18	7	2
2	МПНР	ПММ-03	2022-06-18	7	8
3	ОППРНПО	ПММ-03	2022-06-18	7	10
4	ПИИМРБРСУ	ПММ-03	2022-06-18	7	8

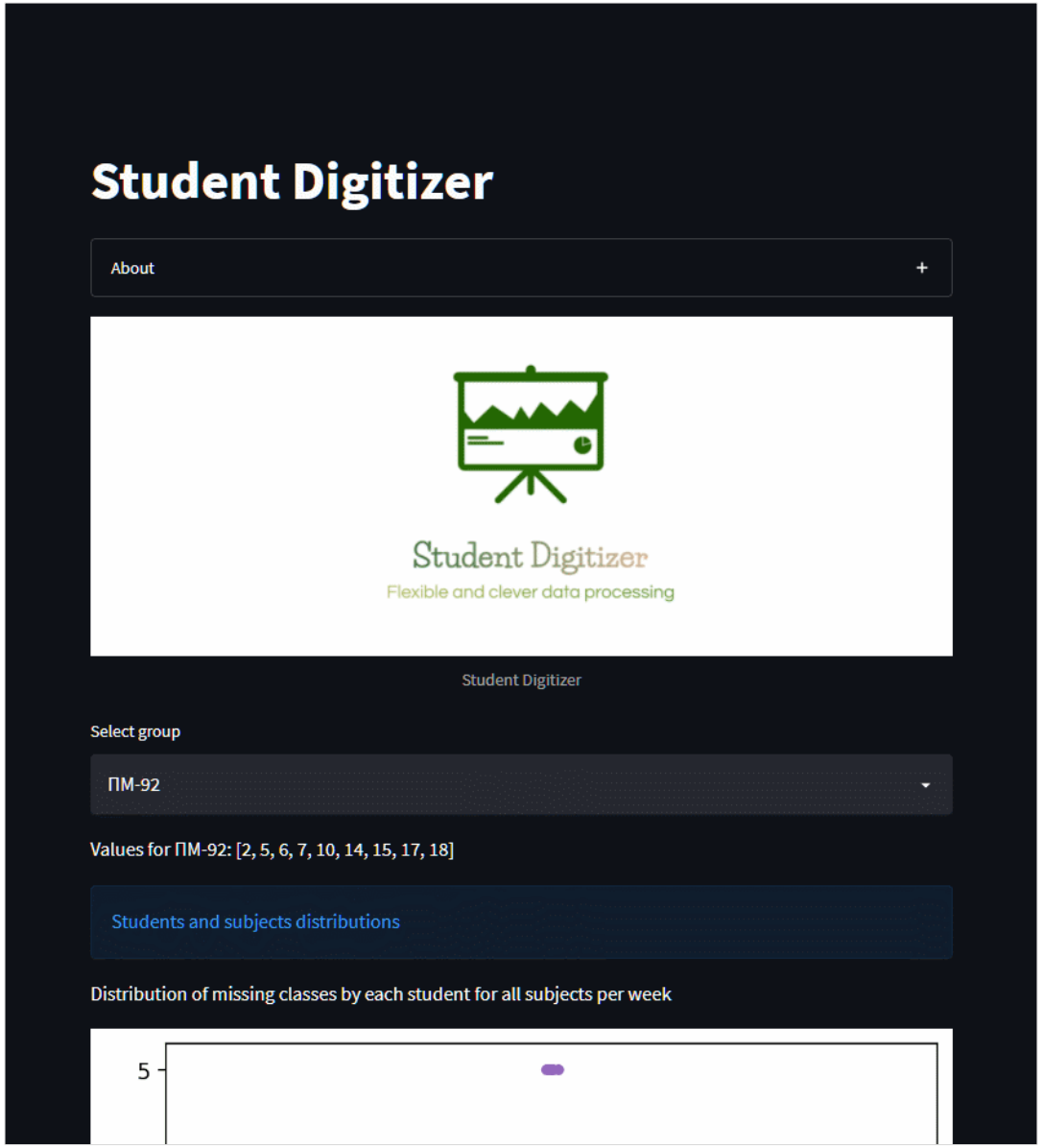
ПАНЕЛЬ ВИЗУАЛИЗАЦИИ ДАННЫХ (ДАШБОРД)

Визуализация полученных результатов в виде информационной панели данных.

GIF файл: https://github.com/EnterSub/Student-Digitizer



ВЕБ-СЕРВИС ВИЗУАЛИЗАЦИИ ДАННЫХ ПОСЕЩАЕМОСТИ СТУДЕНТОВ



Визуализация выходных данных в виде кроссплатформенного веб-сервиса.

Пользователю предлагается выбрать студенческую группу из списка групп, для которых были загружены данные группового журнала из мобильного приложения “Student Digitizer”. После этого отображаются основные актуальные графики распределения количества пропущенных занятий студентом и общего количества пропусков каждого из предметов выбранной группой студентов.

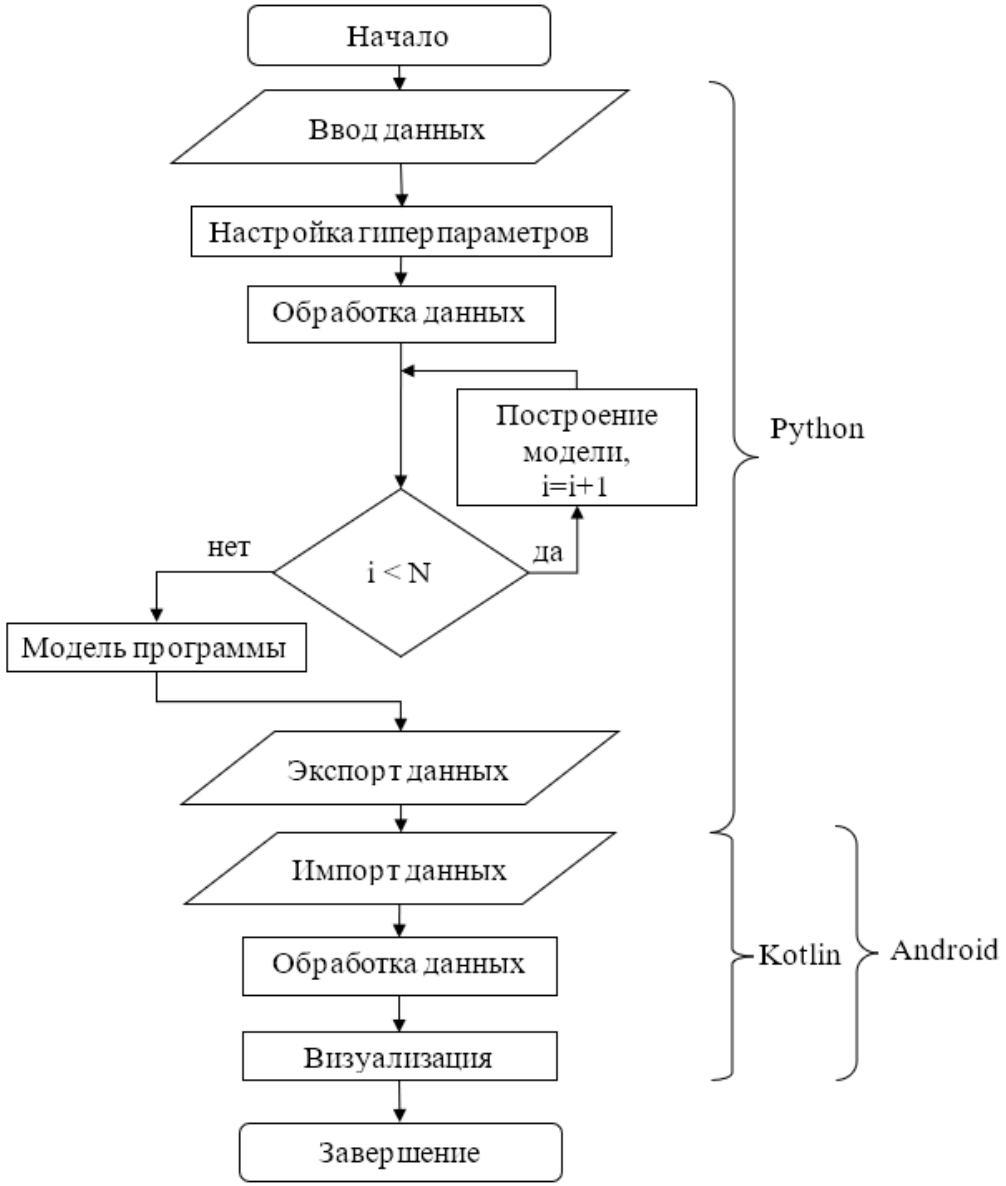
GIF файл: https://github.com/EnterSub/Student-Digitizer

ЗАКЛЮЧЕНИЕ

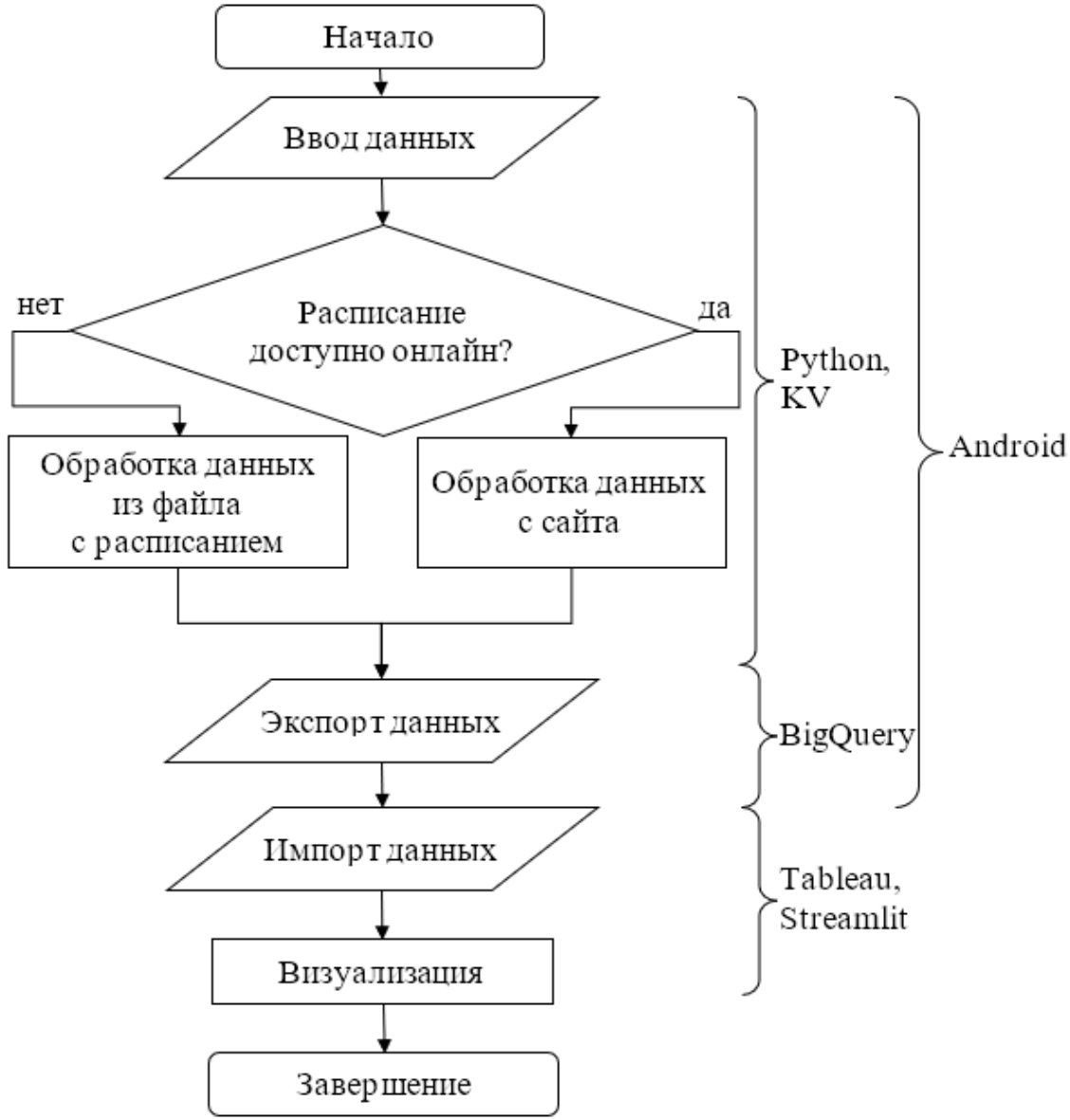
В результате выполнения магистерской диссертации было сделано:

- 2 мобильных приложения, позволяющих классифицировать символы, собирать, обрабатывать и записывать данные в хранилище данных в режиме реального времени;
- 2 типа визуализации данных в виде информационного дашборда и кроссплатформенного веб-сервиса;
- Ускорение визуализации результатов обработки группового журнала после окончания занятия с 1 семестра до 1 недели (~96%).

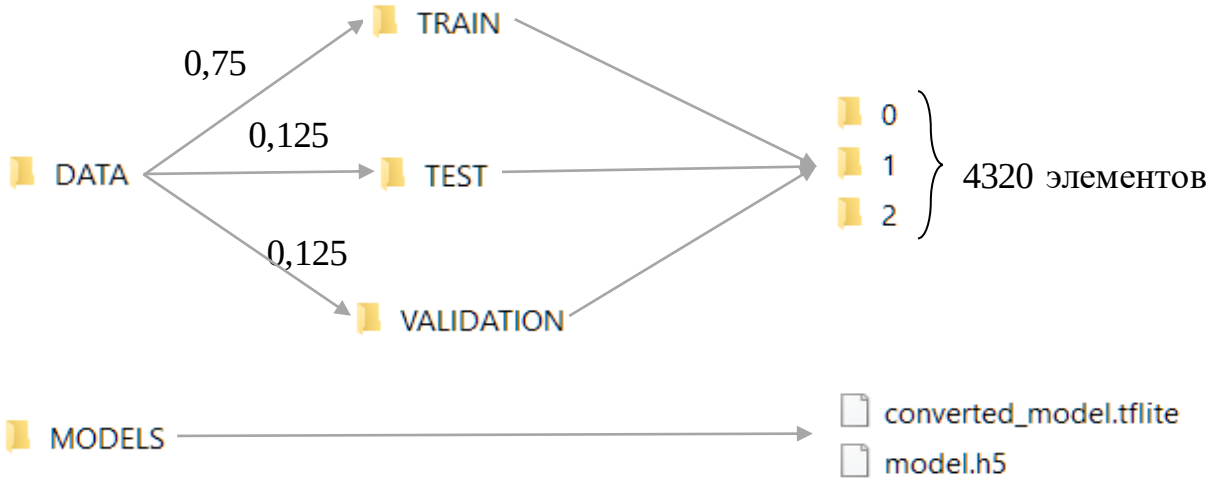
АЛГОРИТМ ПРИЛОЖЕНИЯ “SYMBOL CLASSIFIER”



АЛГОРИТМ ПРИЛОЖЕНИЯ “STUDENT DIGITIZER”



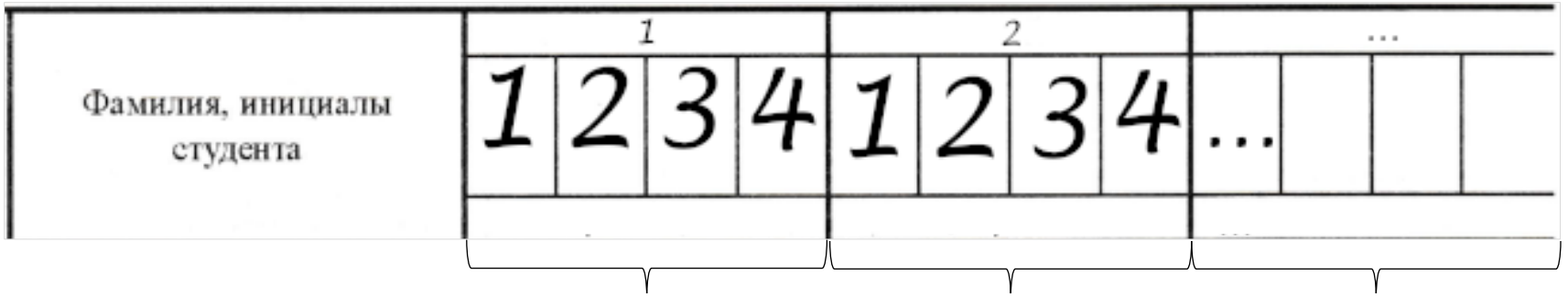
ДАТАСЕТ

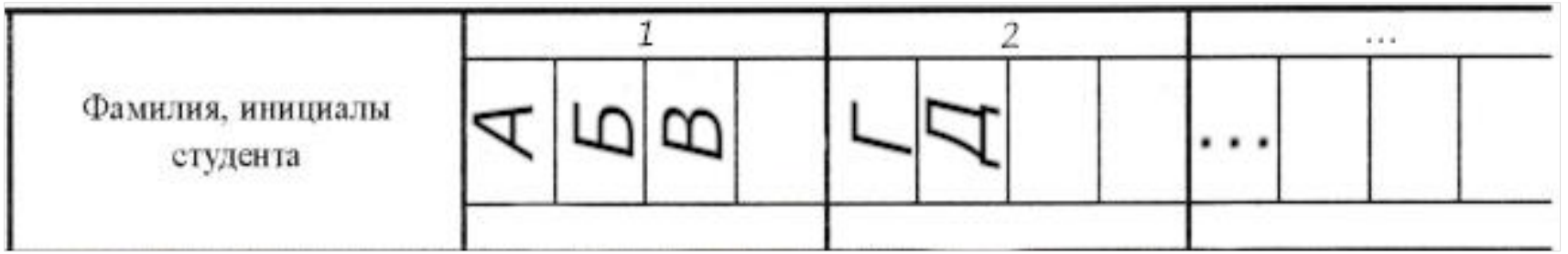

TRAIN
0,75
0,125
DATA
TEST
0,125
VALIDATION
0
1
2
4320 элементов
MODELS
converted_model.tflite
model.h5

ЦИФРОВАЯ ВЕРСИЯ ГРУППОВОГО ЖУРНАЛА

	number	student	column1	column2	column3	column4	column5	column6	column7	column8	column9	column10	colu	ın23	column24	lectures_all
0	ПММ-03			25				26				27				Пропущено часов занят.
1	7	Студент	ИЯ	ПИИМРБРСУ			ОППРНПО	ОППРНПО	КТМИАД		МПНР	МПНР				всего
2		Тип занятия														всего
3	1	Student1	1	1			0	0	0		0	0				2
4	2	Student2	0	1			1	1	1		0	1				5
5	3	Student3	1	0			0	0	0		1	0				2
6	4	Student4	0	1			1	0	0		0	0				2
7	5	Student5	0	0			1	0	0		0	1				2
8	6	Student6	0	1			0	1	0		0	0				2
9	7	Student7	0	1			0	0	0		0	1				2
10	8	Student8	0	0			0	1	0		0	0				1
11	9	Student9	0	1			0	1	0		0	1				3
12	10	Student10	1	0			0	0	0		0	0				1
13	11	Student11	0	0			0	0	0		1	0				1
14	12	Student12	0	1			1	0	0		0	0				2
15	13	Student13	0	0			0	0	1		1	0				2
16	14	Student14	1	0			0	0	0		0	0				1
17	15	Student15	0	0			0	0	0		0	0				0
18	16	Student16	0	1			0	1	0		0	0				2
19	17	Student17	0	0			0	0	0		0	0				0
20	18	Student18	1	0			0	0	0		1	0				2
21	19	Student19	0	0			1	0	0		0	0				1
22	20	Student20	1	0			0	0	0		0	0				1
23																
24																
25																
26																
27																
28																
29																
30																
31																
32																
33																
34																
35																
36																
37		Всего отсутствовало	6	8			5	5	2		4	4				34
38		Подпись преподавателя					Yes	Yes								Yes
39		Подпись старосты										Yes				

ОСОБЕННОСТИ ПРЕДМЕТОВ В ГРУППОВОМ ЖУРНАЛЕ





ОТОБРАЖЕНИЕ ПРЕДМЕТОВ ИЗ РАСПИСАНИЯ ЗАНЯТИЙ

	abbreviate	subject
0	ОППРНПО	Объектно-ориентированный подход при разработке наукоемкого программного обеспечения
1	ПИИМРБРСУ	Прямые и итерационные методы решения больших разреженных систем уравнений
2	КТМИАД	Компьютерные технологии моделирования и анализа данных
3	СКТ	Современные компьютерные технологии
4	ИЯ	Иностранный язык
5	МПНР	Методология представления научно-технических результатов

ПРИМЕР РАСПИСАНИЯ С САЙТА НГТУ

ФПМИ · ПМ-92

Сегодня 12 июня 2022, воскресенье **18 учебная неделя**

		Предмет	Аудитория
ПН	08:30-10:00		
	10:15-11:45		
	12:00-13:30		
	14:00-15:30		
	15:45-17:15		
	17:30-19:00		
	19:15-20:45		
	21:00-22:30		
ВТ	08:30-10:00		
	10:15-11:45		
	12:00-13:30	ПО ЧЁТНЫМ Основы экономических знаний · Крупчатникова В. В. · Практика	2-518
	14:00-15:30	ПО ЧЁТНЫМ Методы оптимизации · Тракимус Ю. В. · Практика	1-2086
		ПО НЕЧЁТНЫМ Методы оптимизации · Филиппова Е. В. · Лабораторная	1-204
	15:45-17:15	Уравнения математической физики · Задорожный А. Г., Персова М. Г., Патрушев И. И. · Практика	1-203а, 1-2036
	17:30-19:00		
	19:15-20:45		
	21:00-22:30		

ПРИМЕР РАСПИСАНИЯ ИЗ HTML ФАЙЛА

ФПМИ · ПМ-92

Сегодня 3 ноября 2021, среда 10 учебная неделя
Предмет
Аудитория
пн
08:30-10:00

10:15-11:45
по чётным Введение в искусственный интеллект и логическое программирование · Авдеенко Т. В. Лабораторная 1-203а
1-203а
по нечётным Операционные системы, среды и оболочки · Кобылянский В. Г., Сивак М. А. Лабораторная 1-204
1-204
12:00-13:30
по чётным Теория вероятностей и математическая статистика · Чимитова Е. В. Лекция 1-426
1-426
по нечётным Операционные системы, среды и оболочки · Кобылянский В. Г. Лекция 1-426
1-426
14:00-15:30

15:45-17:15

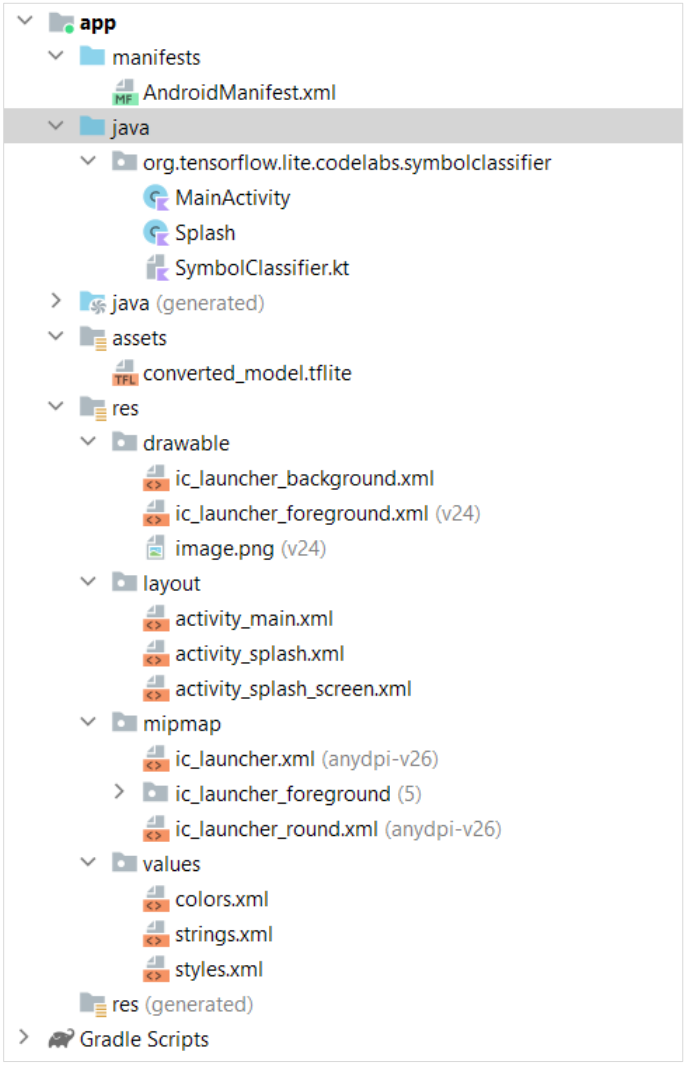
17:30-19:00

19:15-20:45

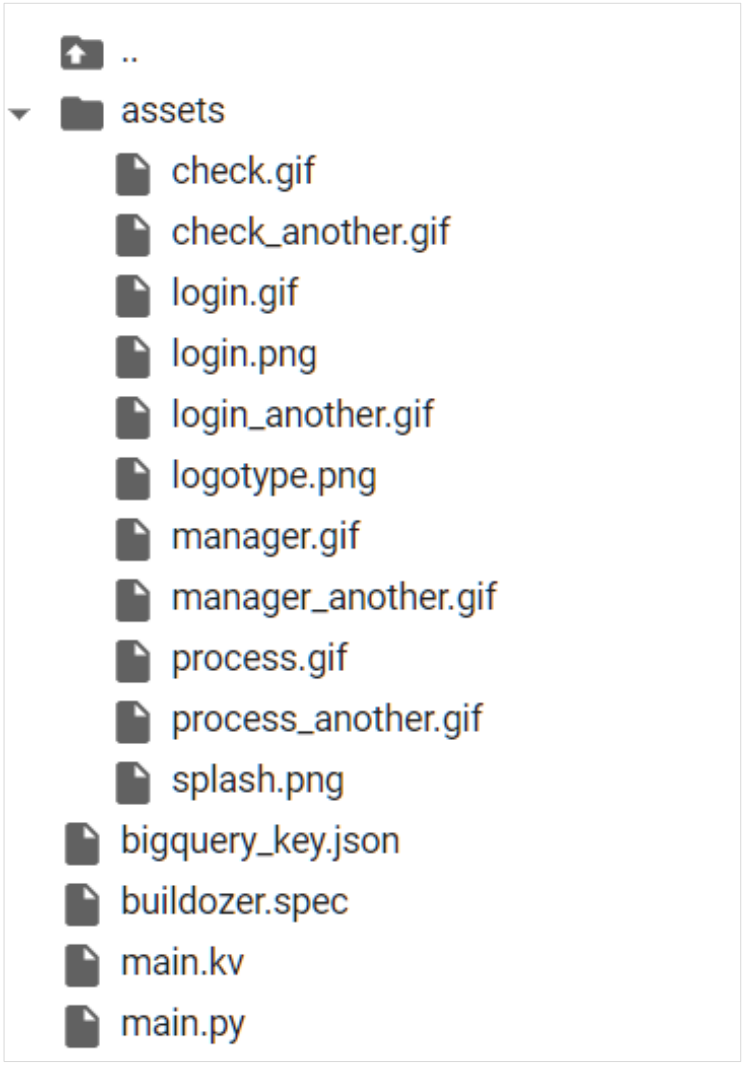
21:00-22:30

СТРУКТУРА МОБИЛЬНЫХ ПРИЛОЖЕНИЙ

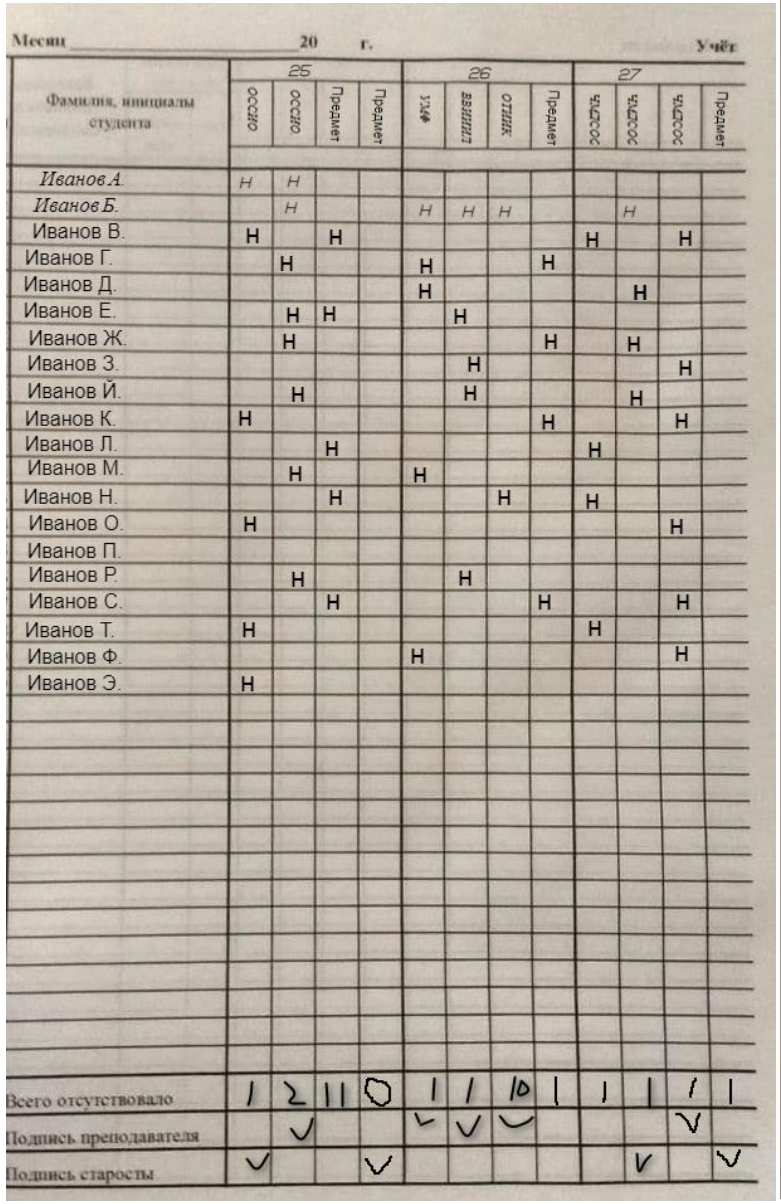
Symbol Classifier



Student Digitizer

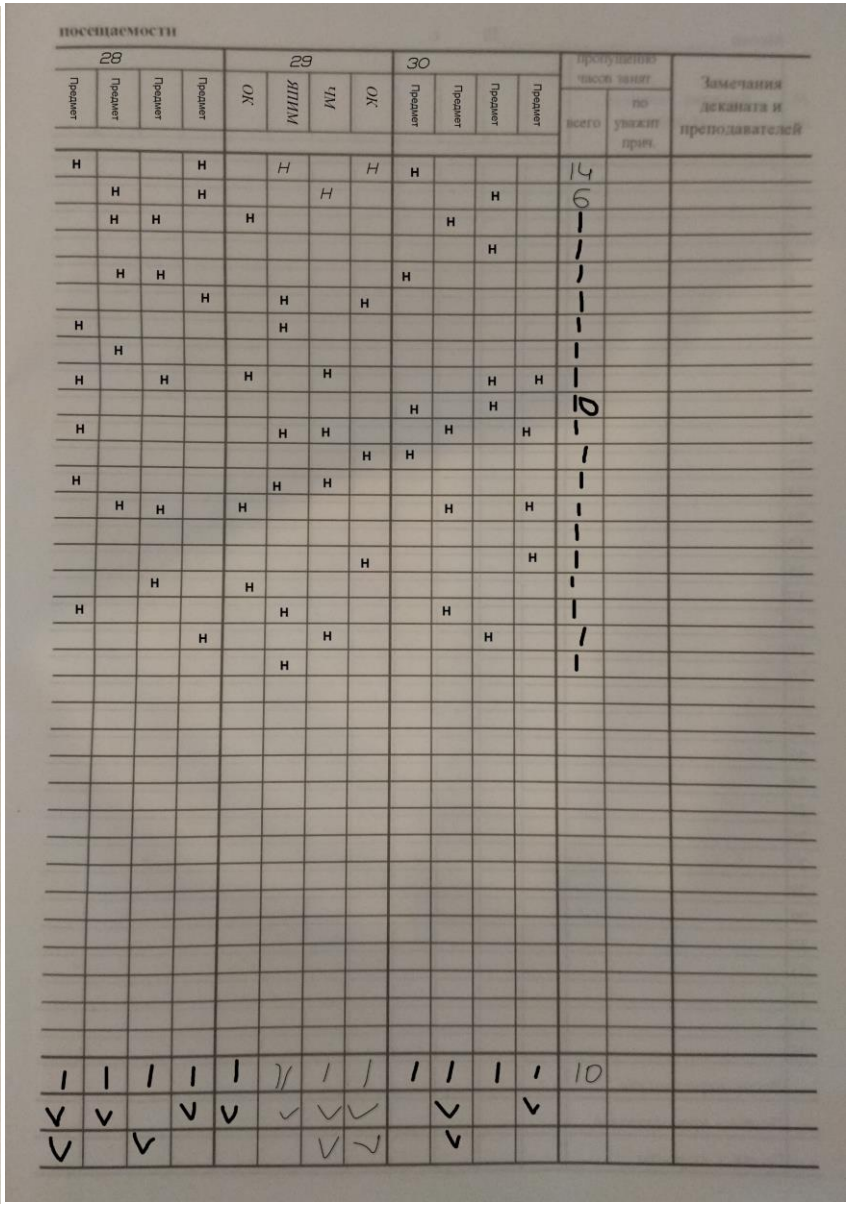


ВХОДНЫЕ ДАННЫЕ



Месяц _____ 20 г. Учёт

Фамилия, инициалы студента	25				26				27			
	ОССЗО	ОССЗО	Предмет	Предмет	УЗФ	ВВППП	ОТППК	Предмет	ЧЗСООС	ЧЗСООС	ЧЗСООС	Предмет
Иванов А.	Н	Н										
Иванов Б.		Н			Н	Н	Н			Н		
Иванов В.	Н		Н						Н		Н	
Иванов Г.		Н			Н			Н				
Иванов Д.					Н					Н		
Иванов Е.		Н	Н			Н						
Иванов Ж.		Н						Н		Н		
Иванов З.						Н					Н	
Иванов Й.		Н				Н				Н		
Иванов К.	Н							Н			Н	
Иванов Л.			Н						Н			
Иванов М.		Н			Н							
Иванов Н.			Н				Н		Н			
Иванов О.	Н										Н	
Иванов П.												
Иванов Р.		Н				Н						
Иванов С.			Н					Н			Н	
Иванов Т.	Н								Н			
Иванов Ф.					Н						Н	
Иванов Э.	Н											
Всего отсутствовало	1	2	11	0	1	1	10	1	1	1	1	1
Подпись преподавателя		✓			✓	✓	✓				✓	
Подпись старосты	✓			✓						✓		✓

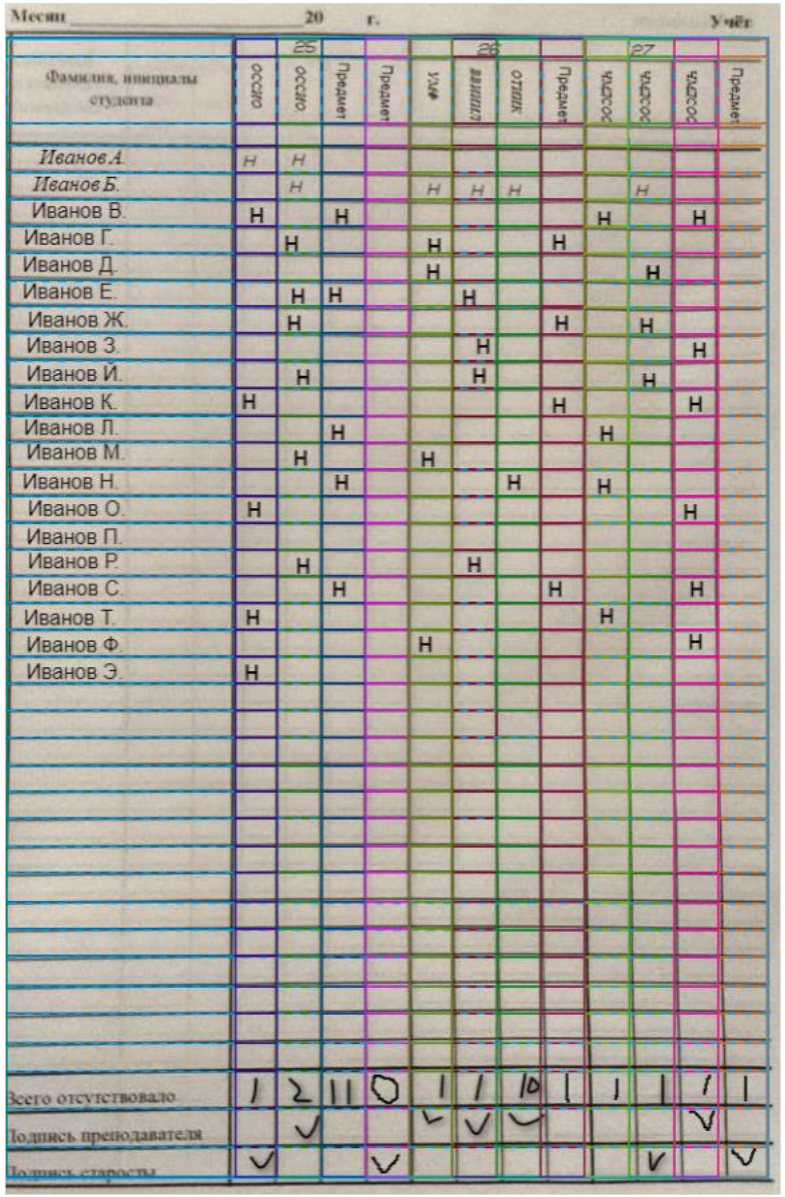


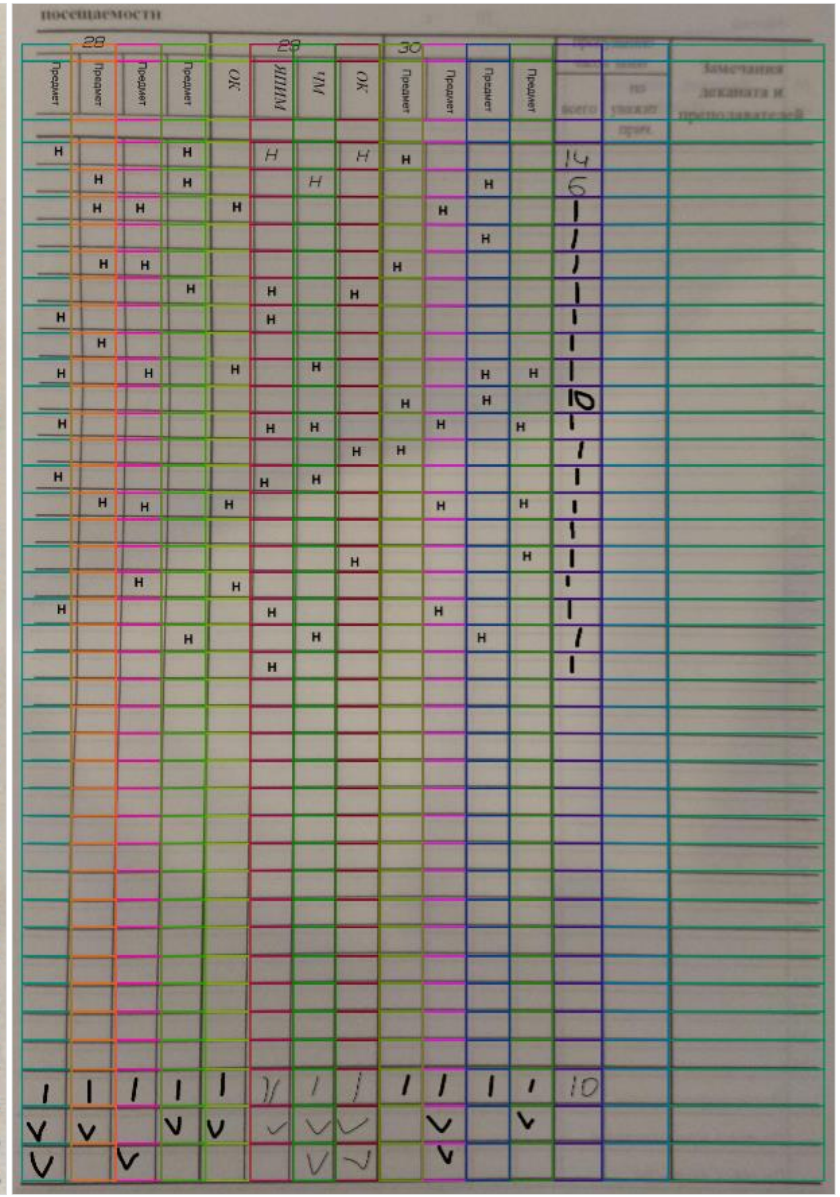
посещаемости

28				29				30				пропущено часов занят.		Замечания деканата и преподавателей
Предмет	Предмет	Предмет	Предмет	ОК	ЯПИМ	ЧМ	ОК	Предмет	Предмет	Предмет	Предмет	всего	по уважит. прич.	
1	1	1	1	1	1	1	1	1	1	1	1	10		
✓	✓		✓	✓	✓	✓	✓		✓		✓			
✓		✓				✓	✓		✓					

Данные журнала вымышленные.

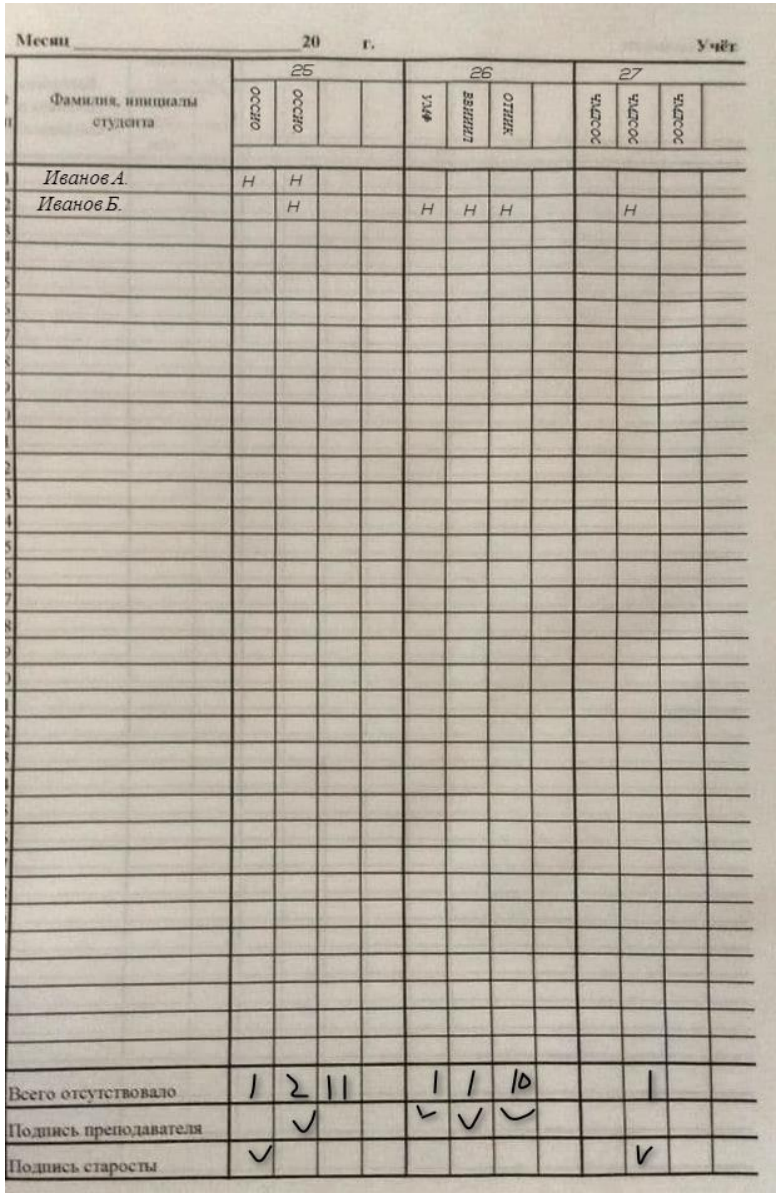
РЕЗУЛЬТАТЫ ОПТИЧЕСКОГО РАСПОЗНАВАНИЯ ТЕКСТА





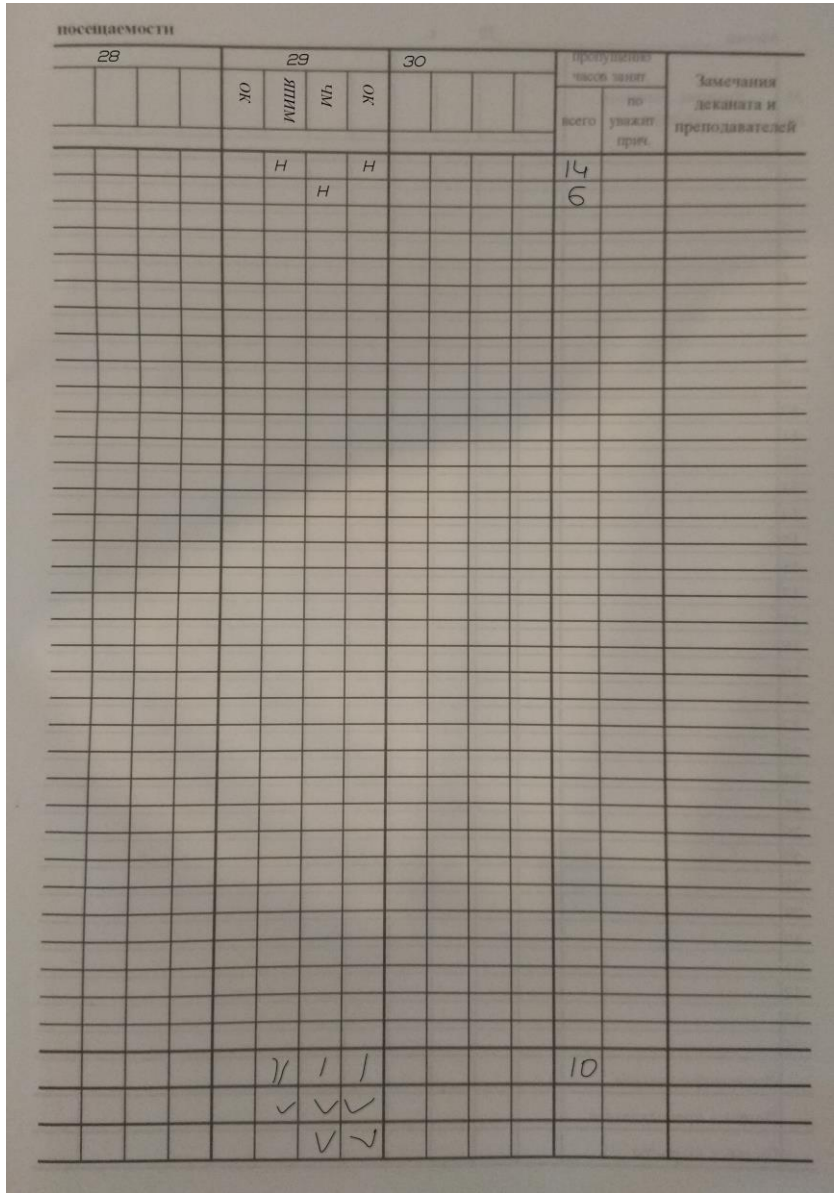
Данные журнала вымышленные.

ВХОДНЫЕ ДАННЫЕ (2 ТЕСТ)



Месяц _____ 20 г. Учёт

Фамилия, инициалы студента	25 ОСНО	25 ОСНО	26 УЛФ	26 ВВИНИТ	26 ОТИНК	27 ЧЫСПОС	27 ЧЫСПОС	27 ЧЫСПОС
Иванов А.	Н	Н						
Иванов Б.		Н	Н	Н	Н		Н	
Всего отсутствовало	1	2	1	1	1	10		1
Подпись преподавателя		✓	✓	✓	✓			
Подпись старосты	✓						✓	



посещаемости

28	29 ЖО	29 ЖИНИМ	29 ЧЫ	29 ЖО	30	пропущено часов занят. всего	пропущено часов занят. по уважит прич.	Замечания деканата и преподавателей
	Н		Н			14		
		Н				6		
	1/	1	1			10		
	✓	✓	✓					
		✓	✓					

Данные журнала вымышленные.

АРХИТЕКТУРА НЕЙРОННОЙ СЕТИ