Министерство науки и высшего образования Российской Федерации

Федеральное государственное автономное образовательное учреждение

высшего образования

**«Пермский национальный исследовательский**

**политехнический университет»**

**(ПНИПУ)**

**ОТЧЕТ**

**о прохождении стажировки**

Севханов Улви Тофиг оглы

(фамилия, имя, отчество)

Интеллектуальный анализ данных и технологии машинного обучения для цифровой трансформации деятельности компании

(наименование программы ДПО)

**Проверил:**

Руководитель ДПП \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(оценка) (подпись) (инициалы, фамилия)

**Пермь 2024**

**ВВЕДЕНИЕ**

1. Место стажировки – компания Databriz. Данная компания занимается разработкой информационно-аналитических систем.

2. Цели стажировки: получение знаний и навыков по подготовке данных для дальнейшего обучения нейросети.

|  |  |  |
| --- | --- | --- |
| Запланированные мероприятия | Сроки | Отчет о выполнении |
| 2 | 3 | 4 |
| Подготовка данных (датасета) | 20.03.2024 - 08.04.2024 | Выполнена |
|  |  |  |

**ОСНОВНАЯ ЧАСТЬ**

1. Программа стажировки состояла в выполнении разметки объектов в заданном наборе данных. В этом наборе данных необходимо было размечать башмаки (bashmak), пирамиды (piramida), колеса поездов (koleso), вентили (valve), коробки с песком (sand\_box) и зелёные/оранжевые жилеты на работниках РЖД (orange\_vest). Каждый объект имеет свой ID (уникальный порядковый номер).

Для того чтобы выполнить разметку, было необходимо изучить функционал Label studio. Label studio — это гибкая платформа для разметки данных для последующего обучения или валидации моделей ИИ.

Label studio позволяет выполнять разметку различными способами. В рамках поставленной задачи было необходимо выполнить разметку областью. Разметка областью подразумевает выделение контура объекта. Притом данный набор данных предполагает выполнение обнаружения объектов, то есть на одном изображении может находиться несколько классов или/и несколько объектов одного класса. Каждой области соответствует свое наименование, которое также иногда называют лейбл(label). Выполнение разметки происходит следующим образом: выбирается класс, далее выполняется расстановка точек, из которых складывается контур. Контуру присваивается соответствующий класс.

Примеры тех классов, которые нужно было разметить на картинках набора данных:



Рисунок 1 — колесо



Рисунок 2 — башмак



Рисунок 3 — пирамида



Рисунок 4 — вентиль



Рисунок 5 — коробка с песком



Рисунок 6 — СИЗ

2. В ходе практики было размечено 500 объектов. Список размеченных объектов указан ниже:

8997

4492

3829

4498

4501

4512

4570

4571

4572

4580

4602

4641

4648

4673

4674

4675

4676

4686

4687

4692

4693

4702

4703

4704

4719

4720

4721

4723

4724

4725

4726

4727

4728

4729

4730

4731

4732

4733

4734

4735

4736

4737

4738

4739

4753

4754

4756

4757

4758

4759

4760

4761

4762

4763

4764

4765

4766

4767

4768

4770

4771

4772

4773

4784

4785

4786

4787

4788

4794

4799

6407

6408

6409

6410

6411

6412

6413

6414

6415

6416

6417

6418

6419

6420

6421

6422

6423

6424

6425

6426

6427

6428

6429

6430

6431

6432

6433

6434

6435

6436

6437

6438

6439

6440

6441

6442

6443

6444

6445

6446

6447

6448

6449

6450

6451

6452

6453

6454

6455

6456

6457

6458

6459

6460

6461

6462

6463

6471

6474

6475

6476

6477

6478

6479

6480

6481

6485

6486

6487

6489

6490

6491

6492

6493

6496

6497

6498

6499

6501

6502

6503

6504

6505

6506

6509

6510

6511

6513

6514

6516

6518

6519

6523

6525

6526

6529

6530

6531

6532

6533

6534

6535

6536

6537

6538

6539

6541

6543

6544

6546

7861

7862

7863

7865

7866

7867

7869

7870

7872

7873

7874

7875

7876

7877

7881

7883

7886

7887

7888

7889

8907

8906

8905

8904

8903

8902

8901

8900

8899

8893

8892

8891

8890

8889

8888

8887

8886

8885

8884

8883

8882

8881

8880

8879

8824

8825

8823

8822

8814

8813

8812

8811

8810

8809

8808

8807

8806

8805

8804

8803

8802

8801

8800

8799

8798

8797

8796

8794

8792

8790

8788

8786

8784

8782

8780

8778

8776

8774

8772

8770

8768

8766

8764

8762

8760

8742

8740

8738

8736

8734

8731

8728

8725

8722

8717

8702

8698

8695

8694

8691

8690

8687

8686

8682

8678

8674

8670

8662

8643

8641

8636

8628

8626

8614

8608

8606

8602

8596

8590

8584

9012

9013

9028

9029

9031

9033

9036

9037

9038

9039

9043

9044

9045

9047

9049

9054

9056

9060

9068

9070

9071

9072

9075

9078

9080

9081

9083

9085

9086

9088

9091

9093

9096

9099

9105

9106

9107

9108

9112

9113

9116

9118

9119

9120

9121

9122

9124

9128

9132

9133

9336

9337

9338

9339

9340

9341

9343

9345

9346

9347

9350

9352

9355

9358

9359

9361

9363

9364

9365

9366

9367

9371

9373

9374

9376

9377

9378

9379

9380

9381

9382

9385

9386

9387

9388

9391

9393

9394

9396

9399

9400

9401

9402

9403

9404

9405

9406

9408

9409

9411

9412

9413

9414

9420

9421

9422

9424

9426

9427

9429

9431

9432

9434

9436

9437

9439

9442

9443

9445

9447

9450

9451

9453

9455

9457

9458

9460

9461

9465

9466

9467

9468

9472

9475

9476

9478

9480

9481

9485

9494

9496

9500

9501

9504

9506

9507

9510

9513

9519

9520

9523

9529

9532

9533

9535

9536

9537

9538

9541

9543

9545

9547

9550

9551

9554

9555

9558

9559

9562

9566

9568

9569

9571

9573

9574

9575

9577

9578

9580

9582

9583

9587

9593

9594

9595

9597

9599

9600

9607

9608

9610

9612

9616

9618

9620

9623

9624

9625

9635

9636

**ЗАКЛЮЧЕНИЕ**

1. Во время выполнения практики были получены необходимые знания и навыки по подготовке данных для дальнейшего обучения нейросети. Получен опыт работы с Label studio.

2. Результаты стажировки в дальнейшем могут быть использованы для обучения нейросети.

Стажер \_\_\_\_\_\_\_\_\_\_\_ Севханов У.Т.

*(подпись стажера)**(расшифровка подписи)*

Дата