МИНИСТЕРСТВО ЦИФРОВОГО РАЗВИТИЯ, СВЯЗИ И МАССОВЫХ

КОММУНИКАЦИЙ РОССИЙСКОЙ ФЕДЕРАЦИИ

Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования

**«Московский Технический Университет Связи И Информатики (MTUCI)»**

Кафедра «Математическая кибернетика и информационные технологии»

**Лабораторная Работа 4**

по дисциплине

«**Машинное обучение**»

Выполнил: студент 3 курса гр. БВТ2201

Ньяти Каелиле

Москва 2025 г

**Содержание**

1. Задание
2. Ход работы
3. Вывод
4. **Задание**

Написать BERT модель которая будет способна классифицировать текст на токсичный или не токсичный. Датасет можно выбрать любой, но если лень искать то можете взять этот. Аккуратно выбирать размеры модели так как у RNN моделей большие проблемы со скоростью обучения. В качестве токенизатора можно взять любой токенизатор с hf либо написать свой эмбеддинг слой который будет превращать именно слова в эмбеддинги, тут по желанию.

В качестве лосса это BCE. В качестве оптимизатор Adam/AdamW, lr=3e-4. Но если захотите что другое то можно другое. Loss можете взвесить так как явный дисбаланс классов. По метрикам считайте accuracy, f1, roc-auc, pr-auc. Не забудьте проводить расчеты на cuda.

1. **Ход работы**

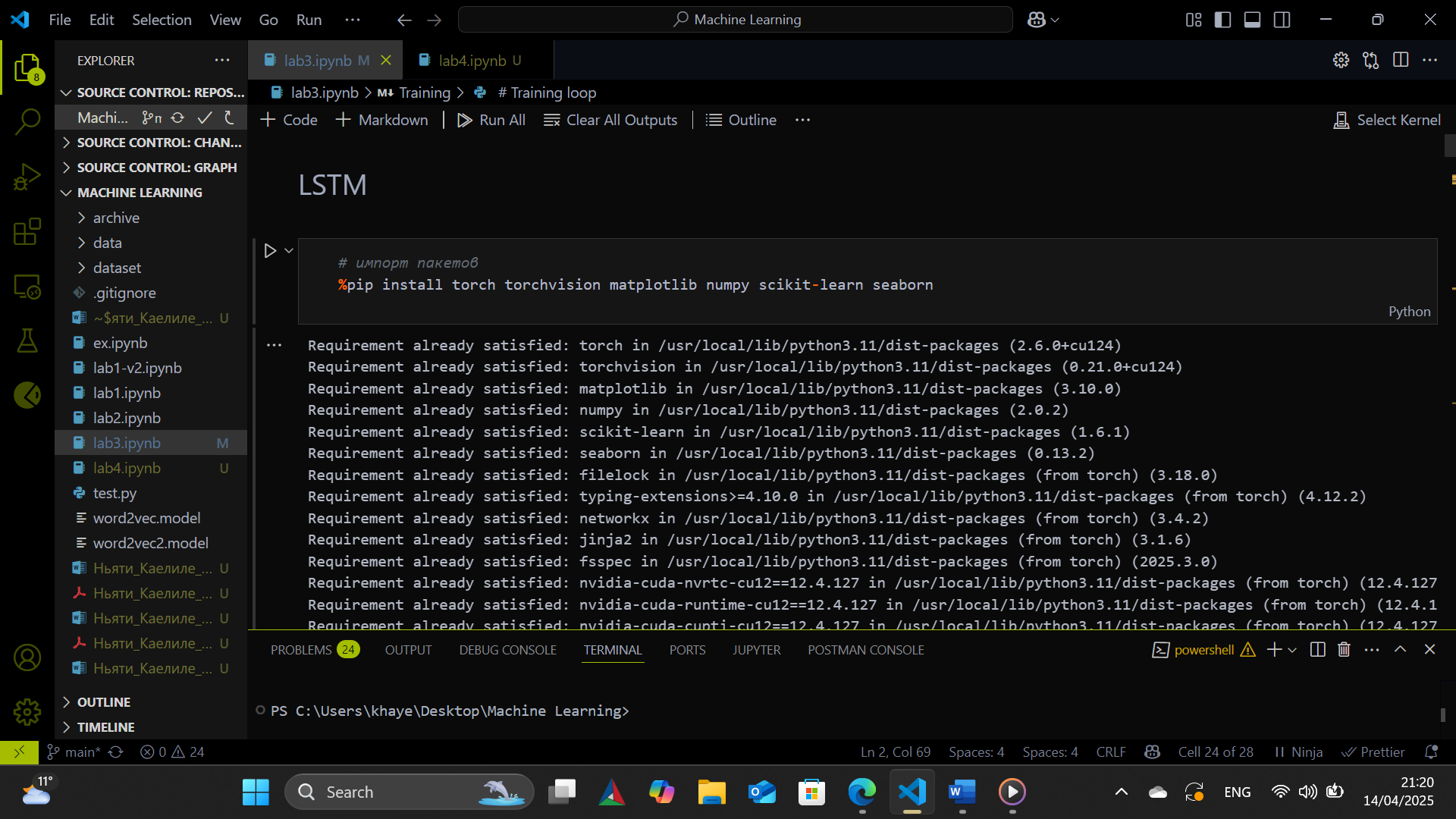


Рис 1. Библиотекы

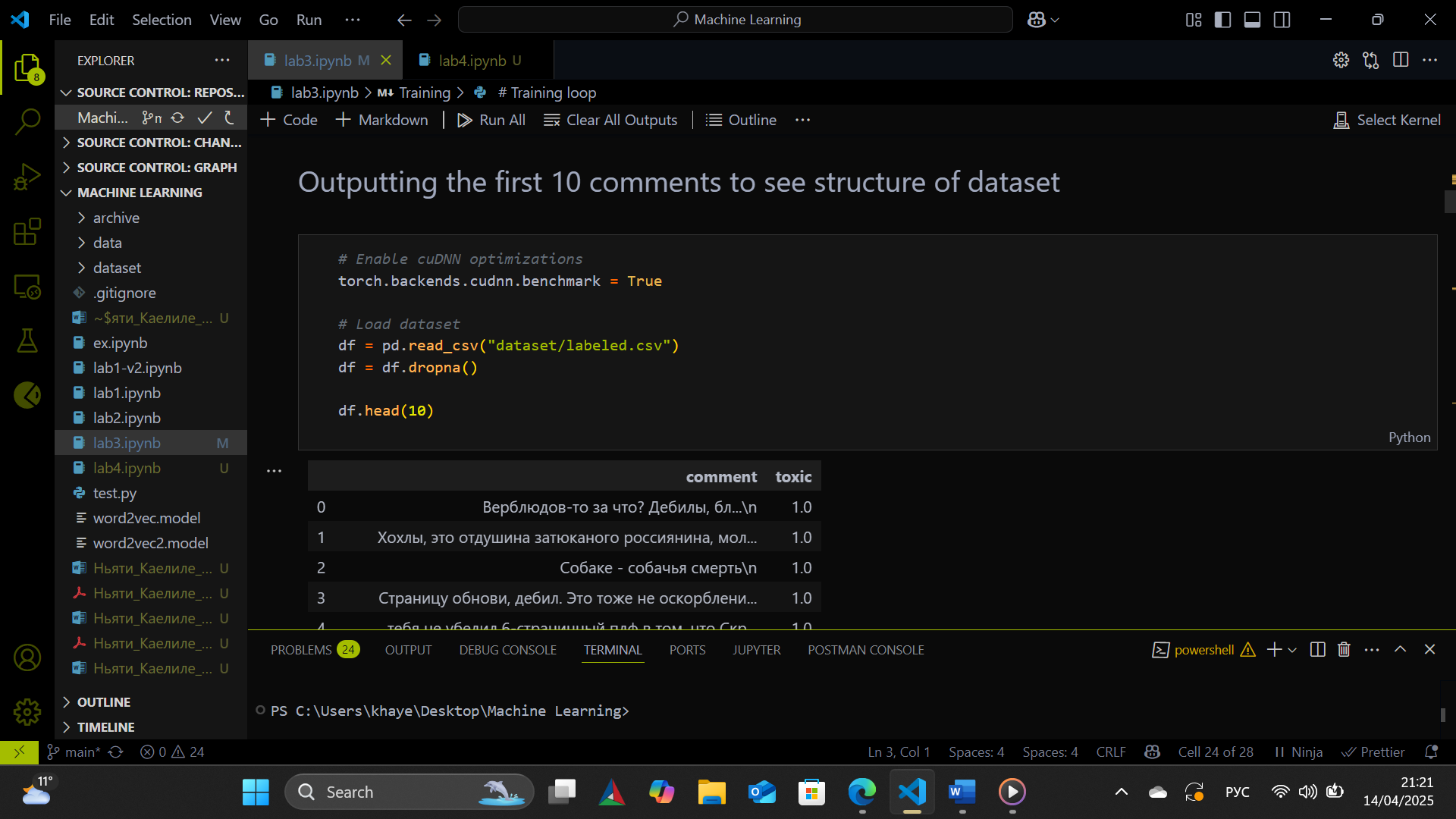
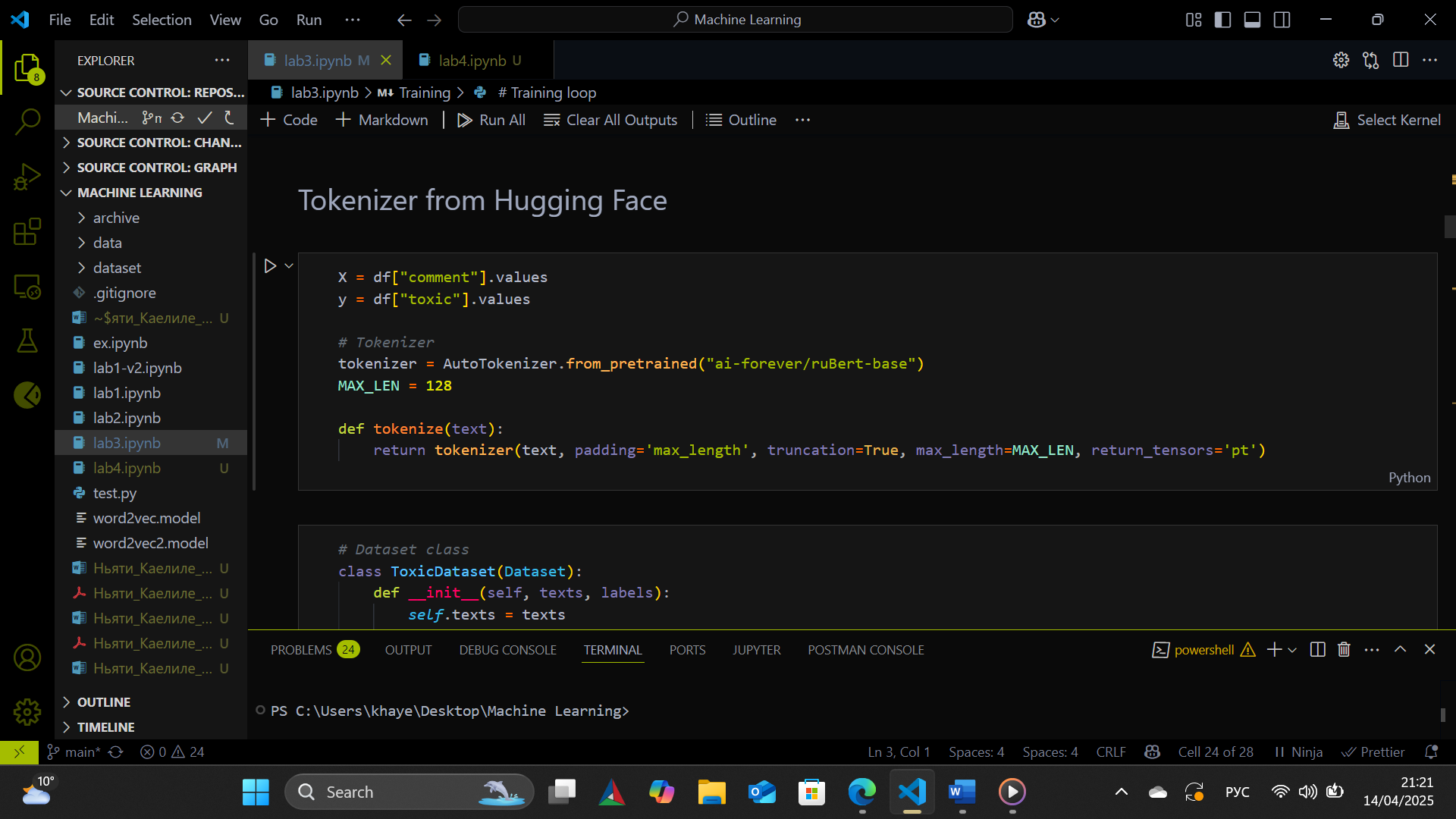


Рис 2. Дата



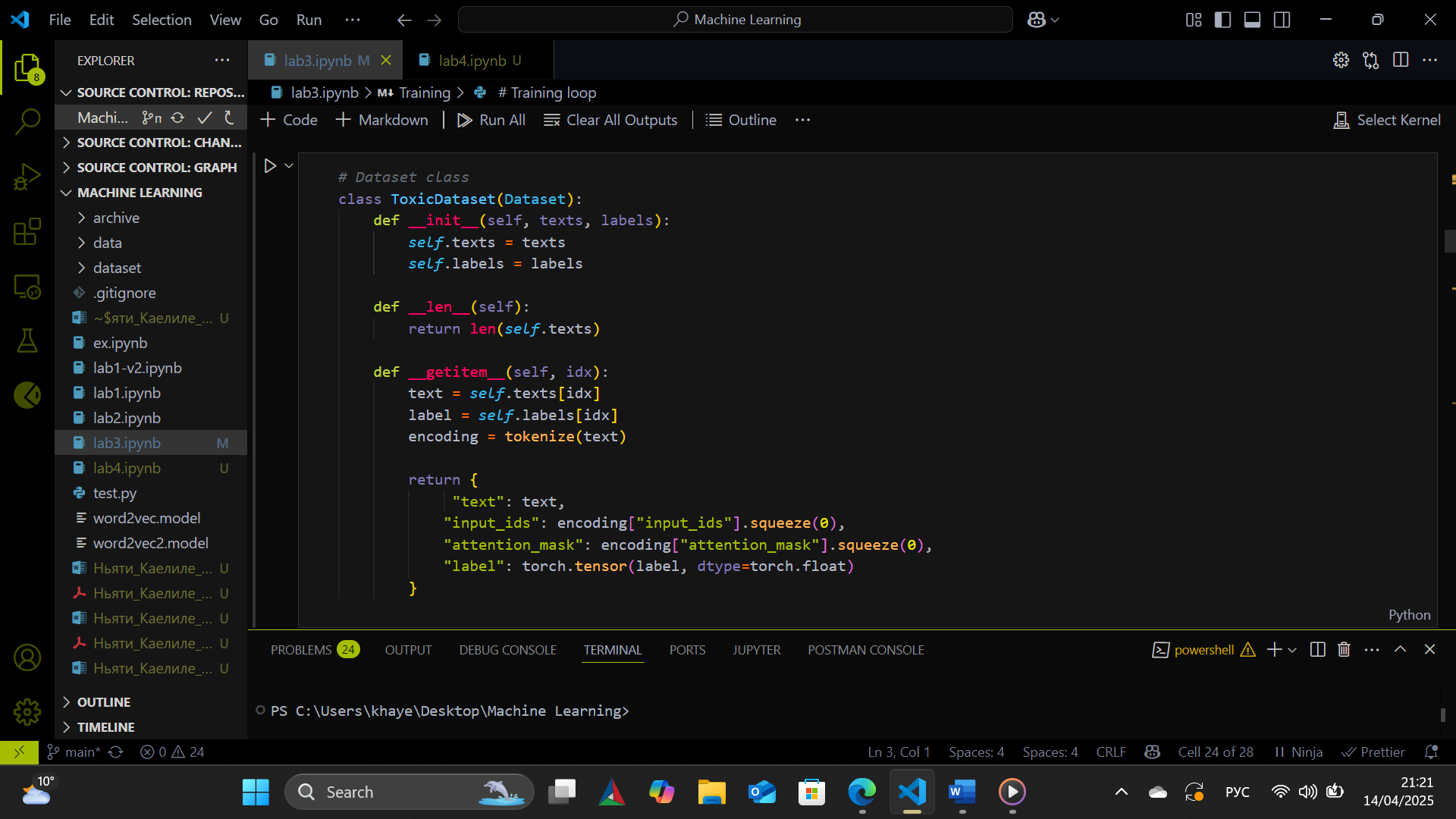
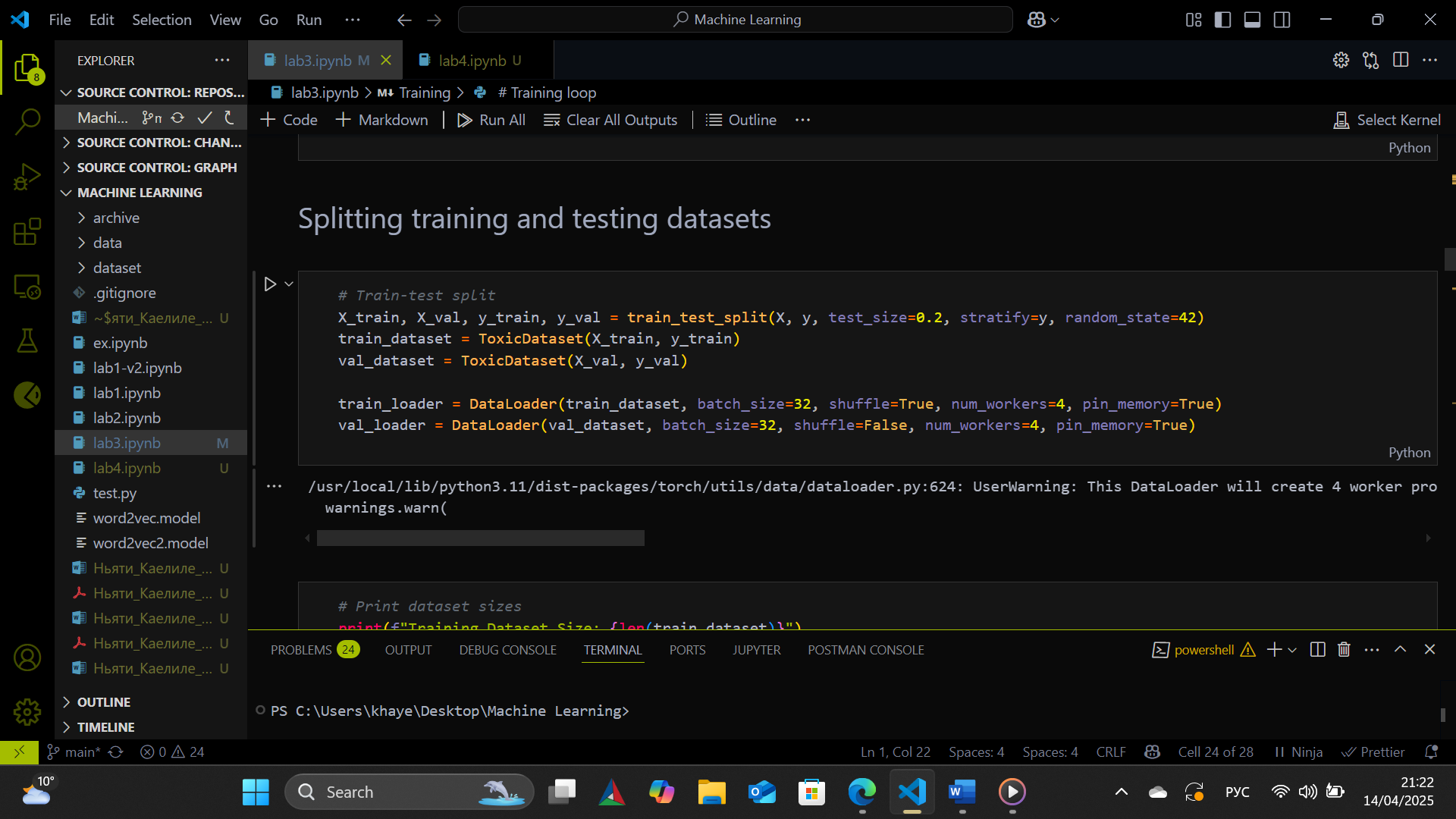
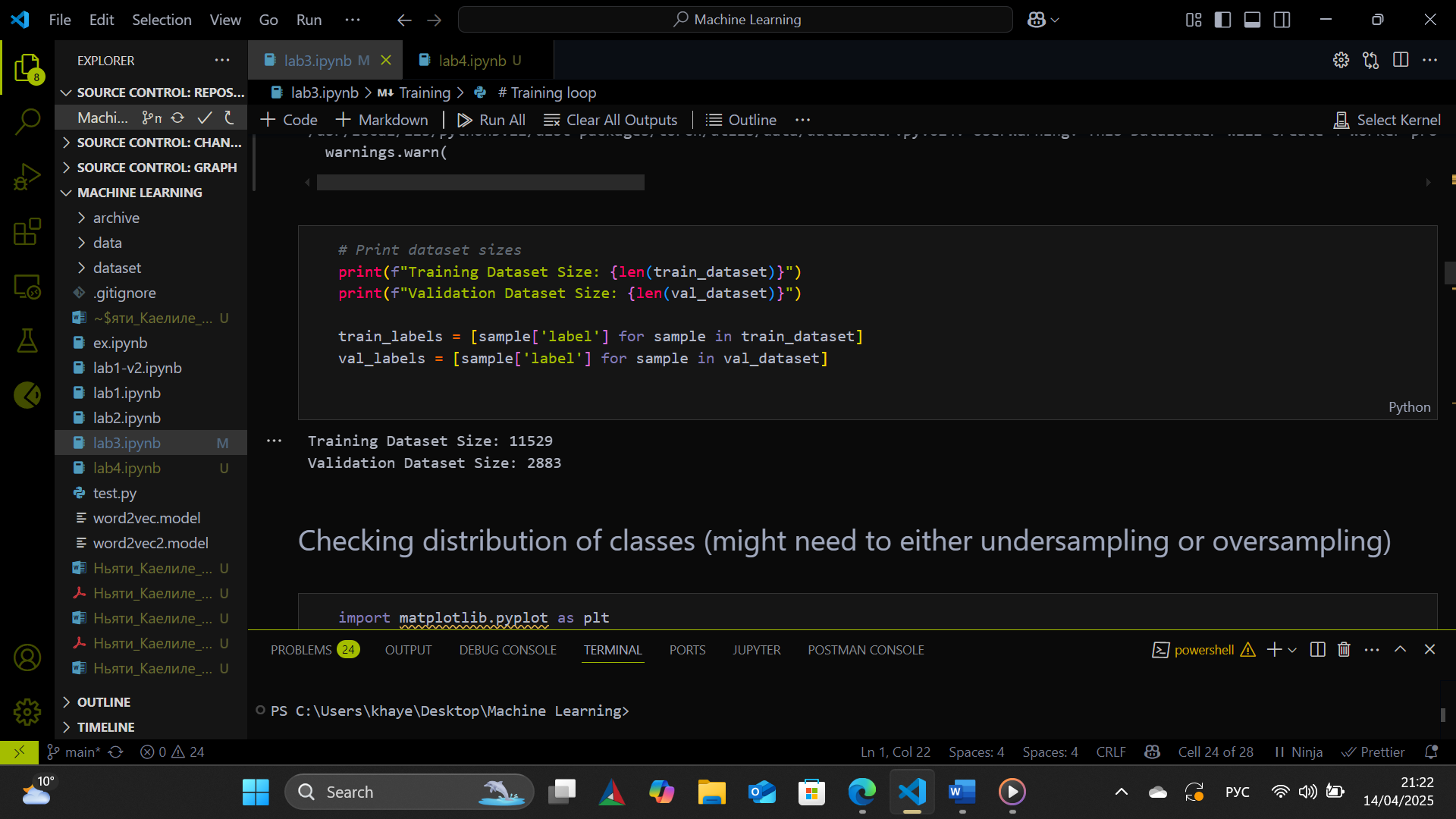
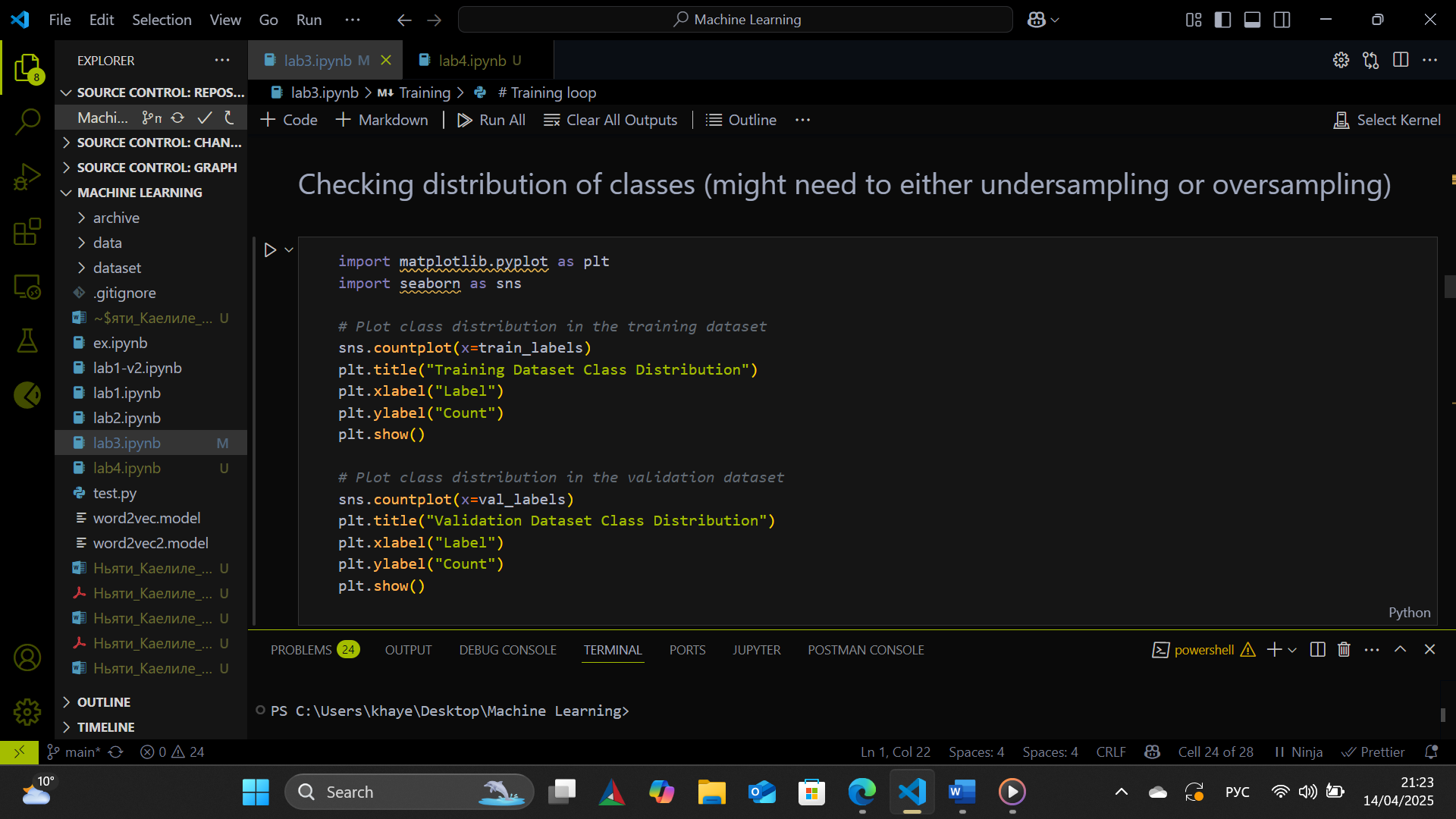
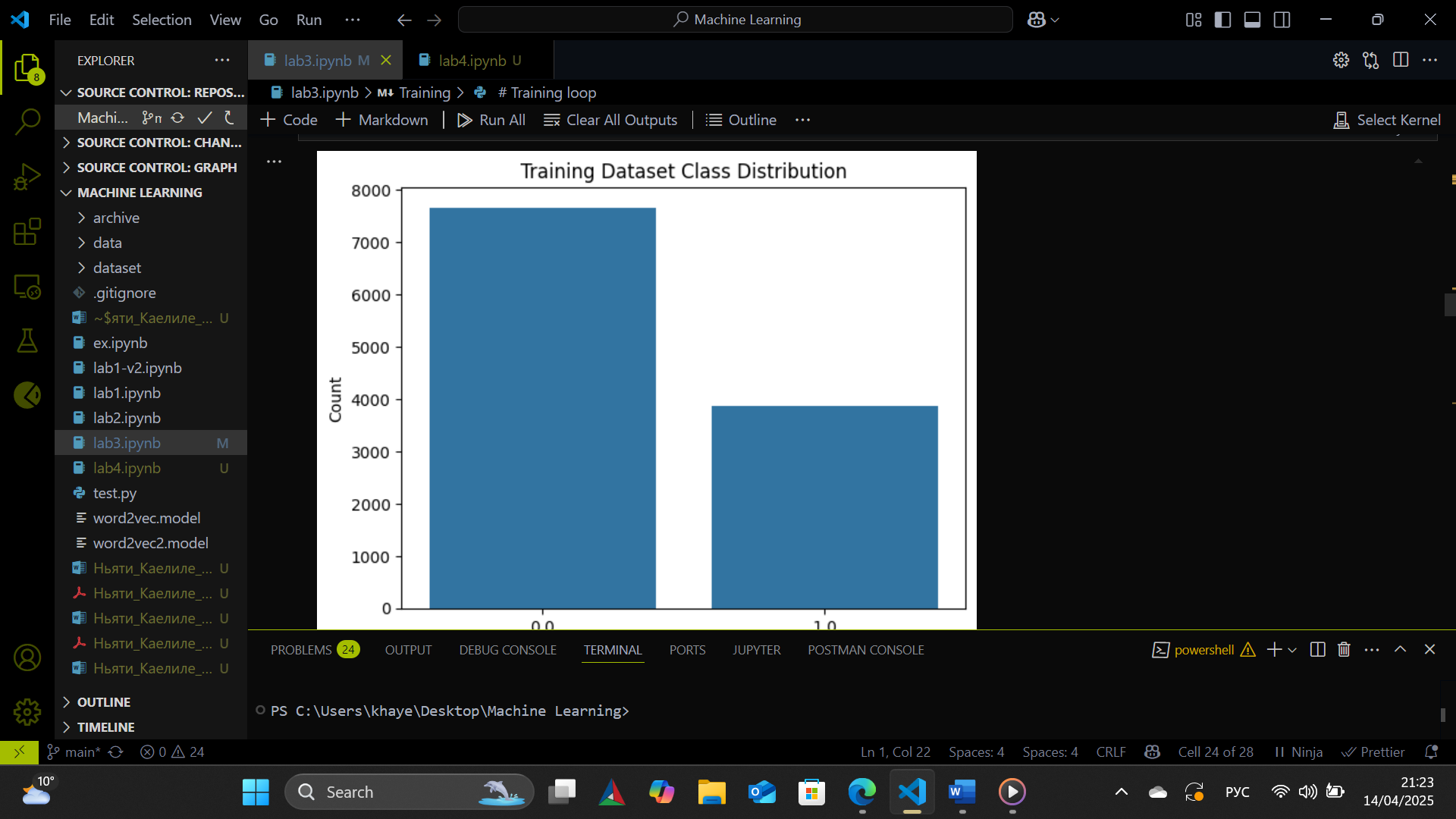


Рис 3. Токенизатор









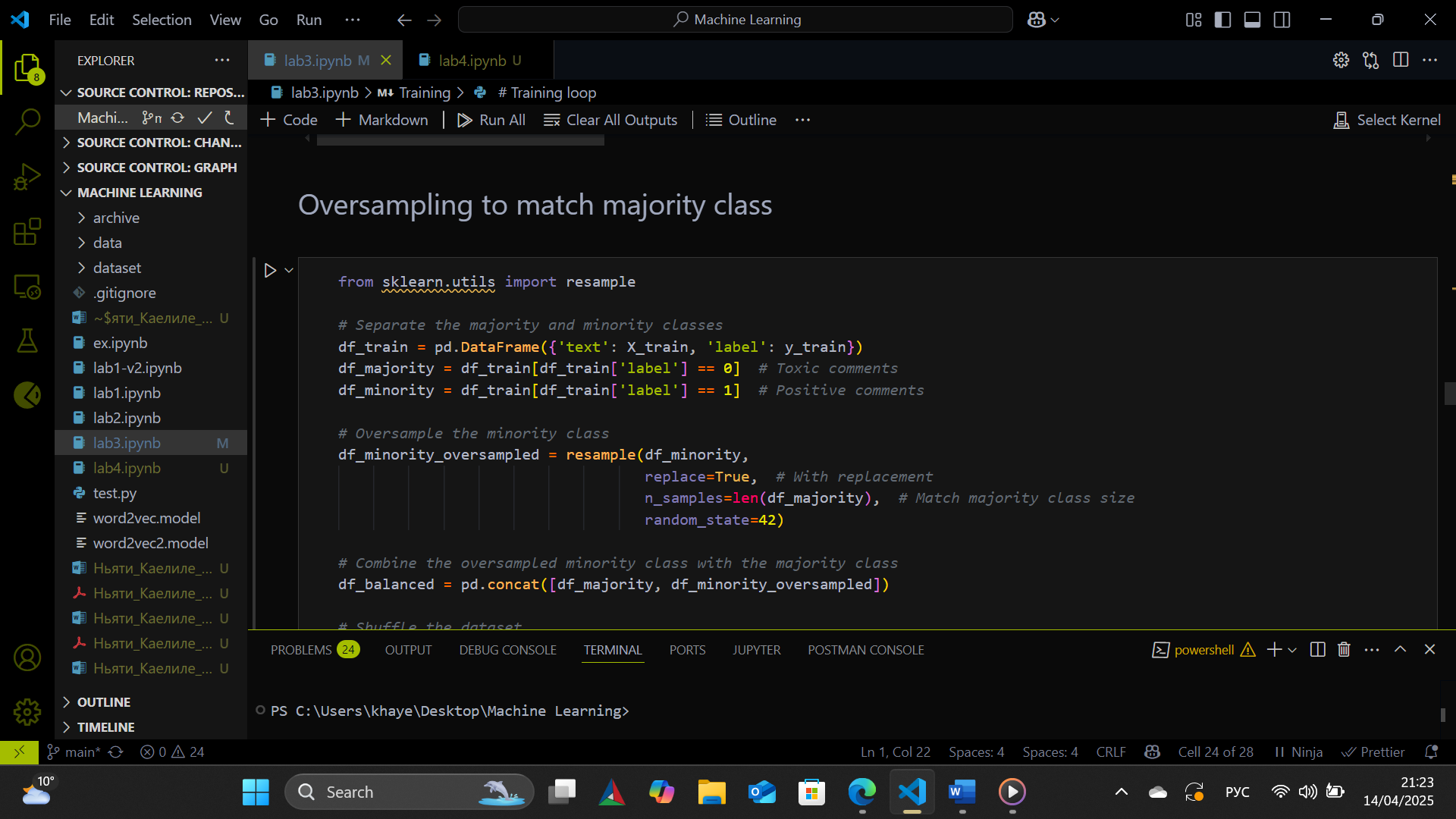
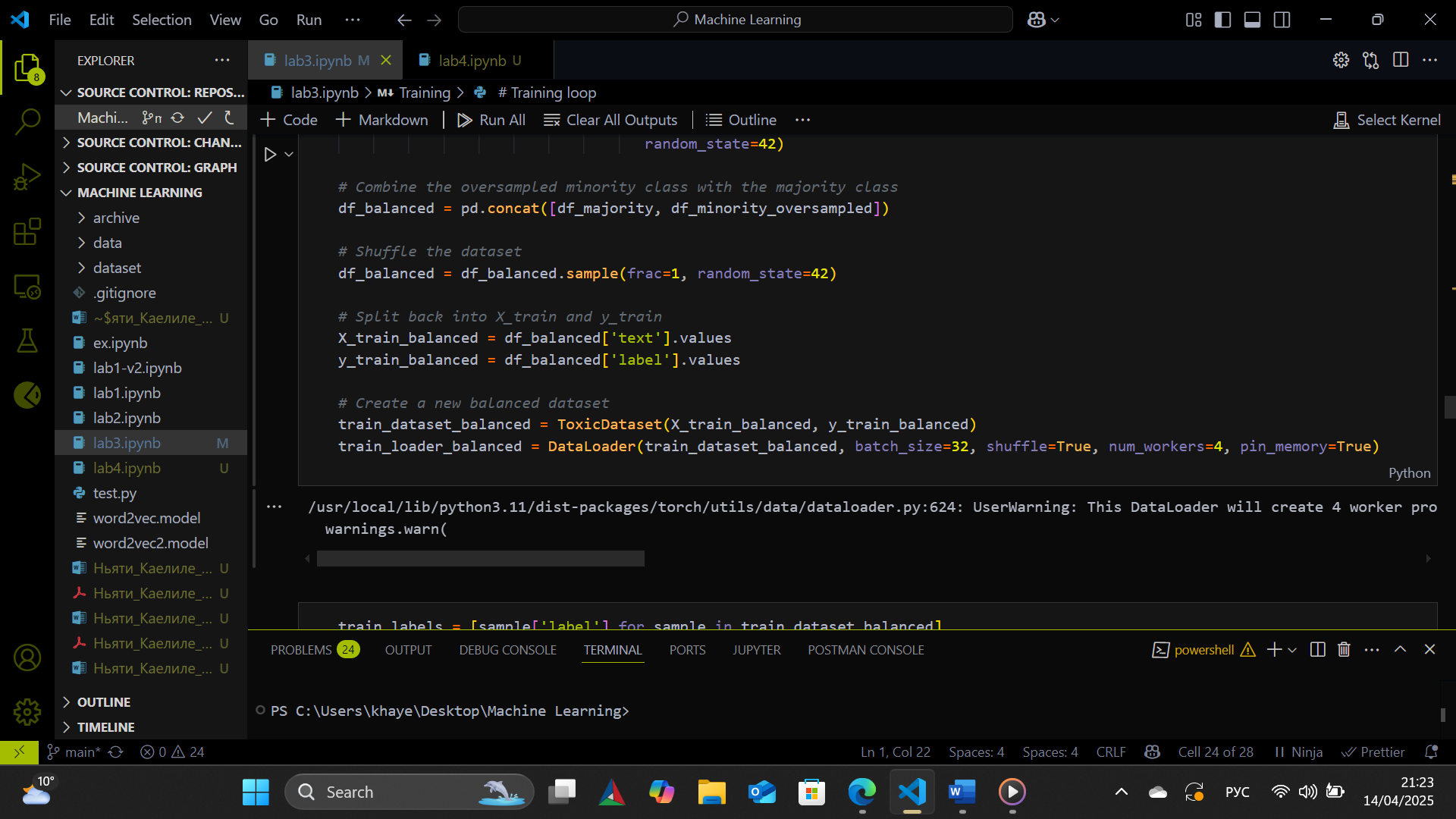
 



Рис 4. Oversampling

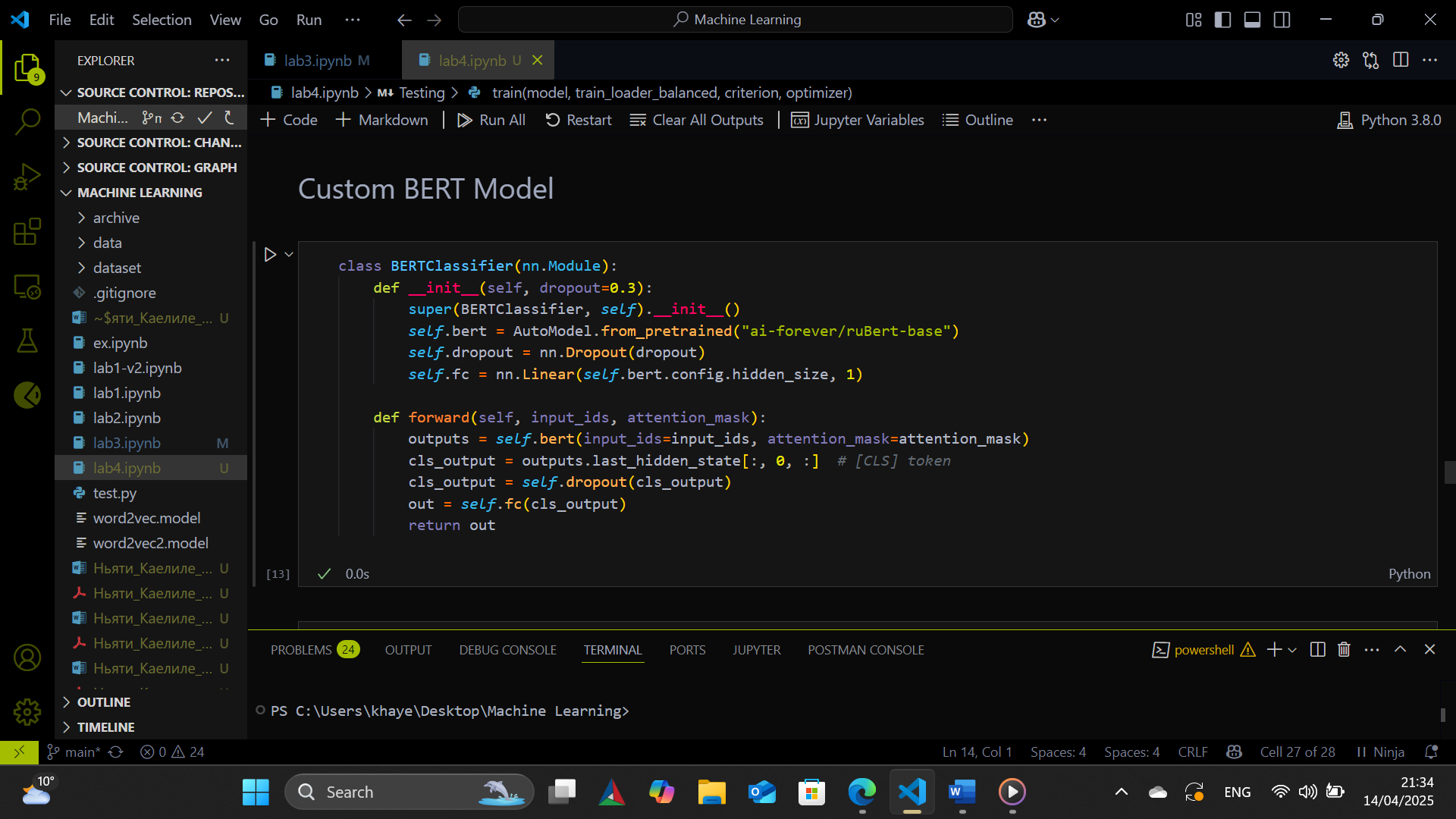
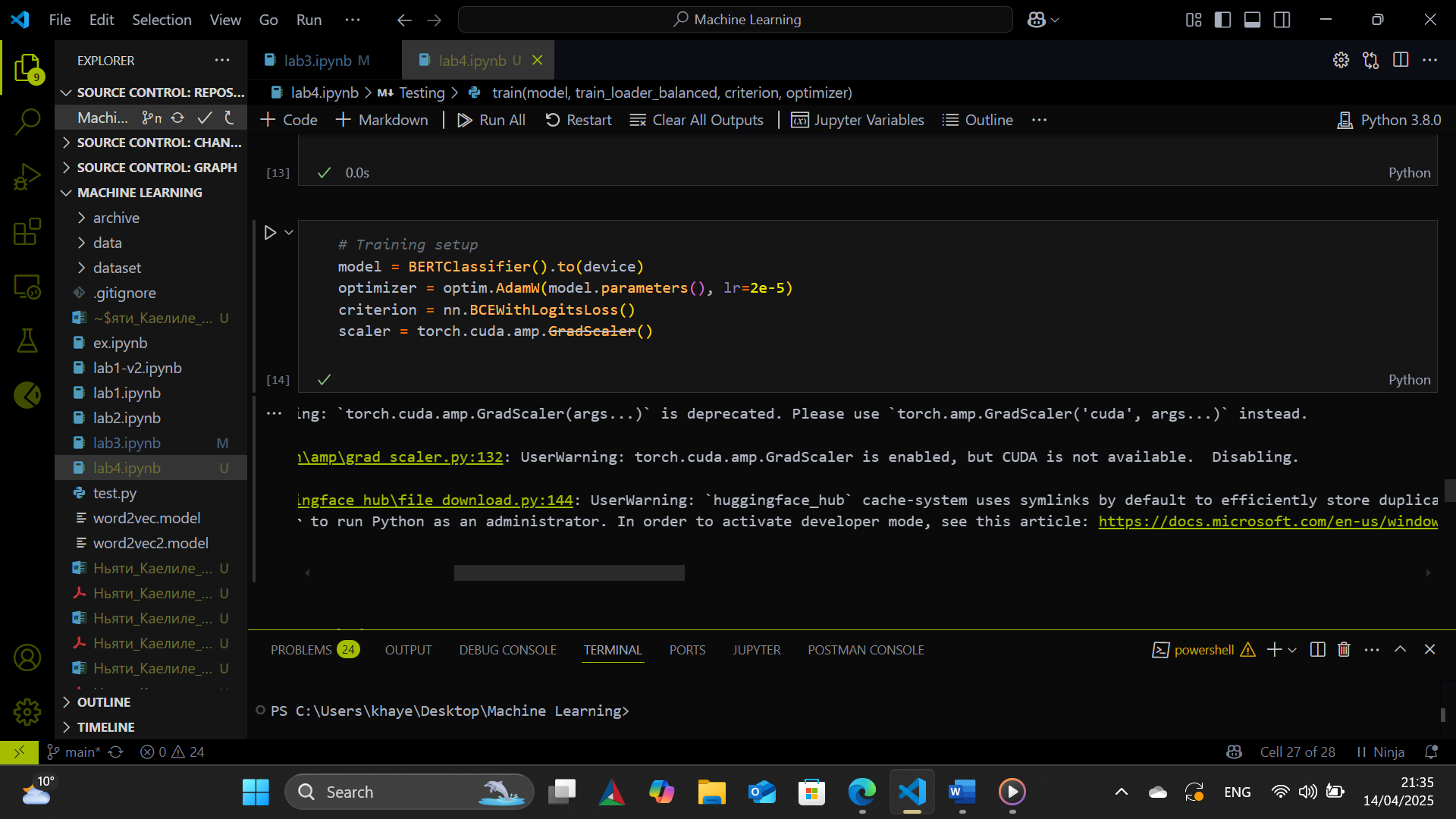
 

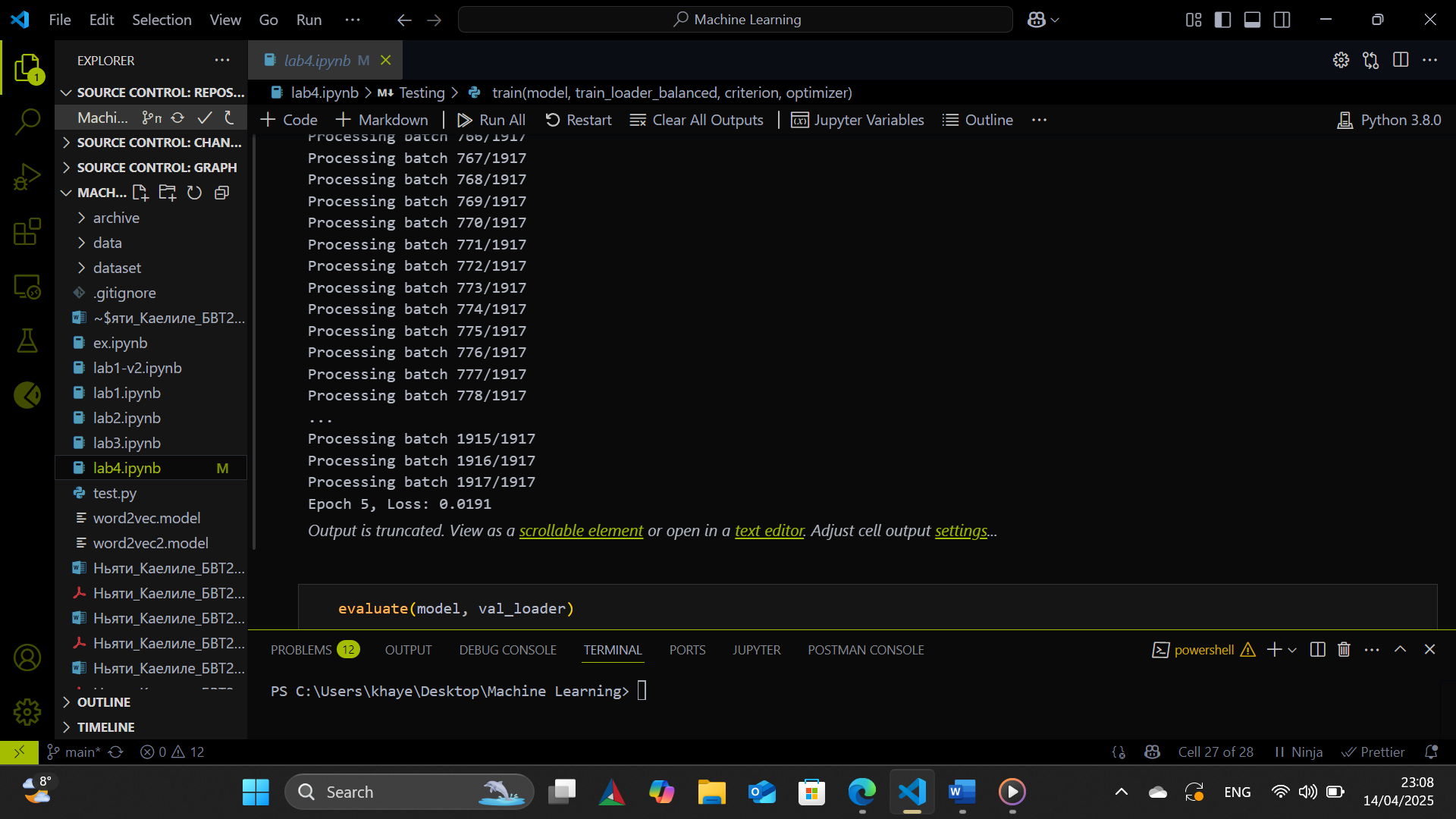
Рис 5. BERT model



Рис 6. Тренирование







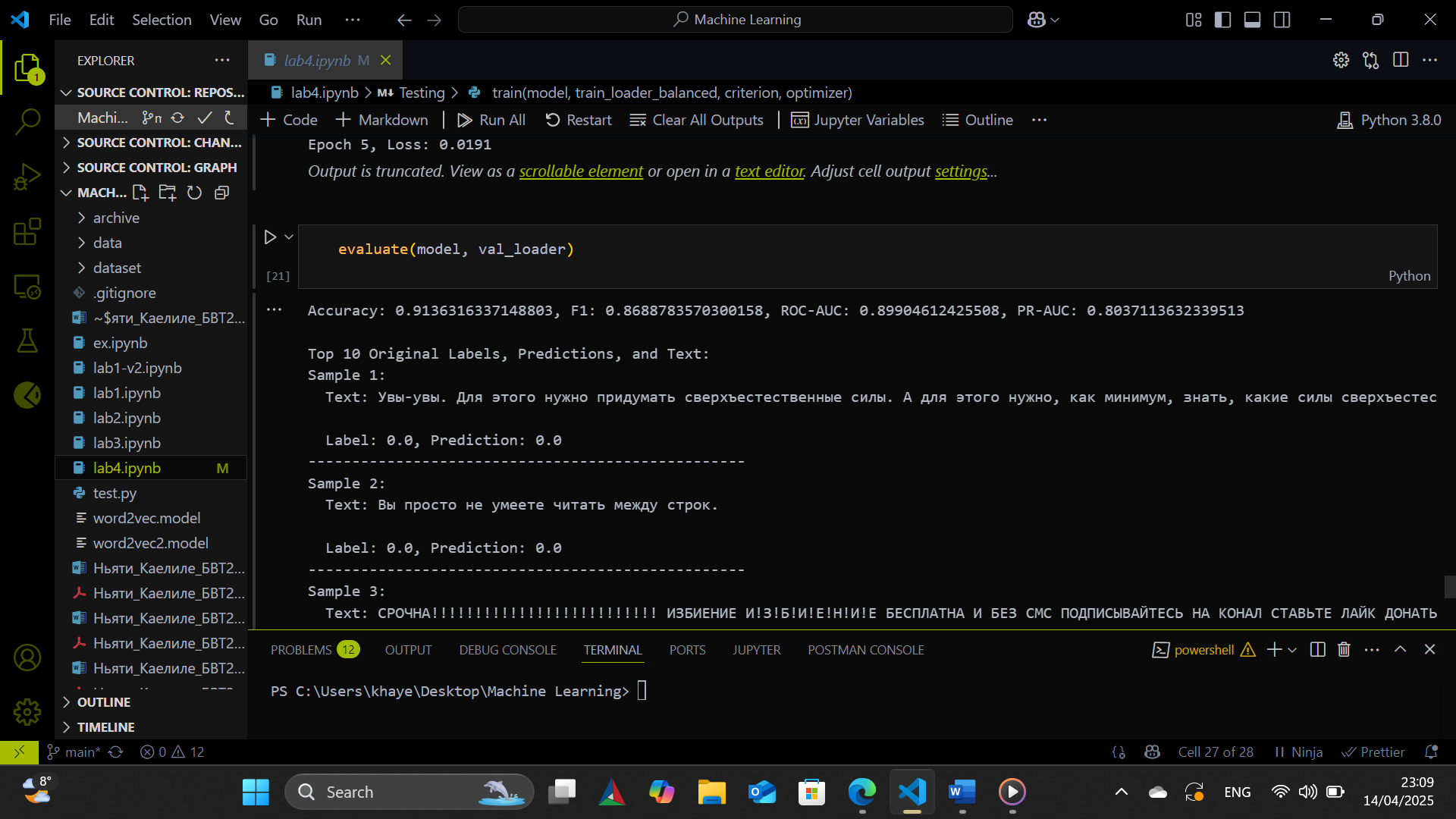


Рис 7. Результаты

1. **Выводы**

В рамках данной задачи была разработана модель на основе BERT для бинарной классификации текста на токсичный и нетоксичный. Благодаря мощным контекстным эмбеддингам BERT и использованию токенизатора из библиотеки Hugging Face, модель получила возможность учитывать семантику и структуру языка при анализе текста. Для борьбы с дисбалансом классов была применена взвешенная функция потерь Binary Cross-Entropy, а оптимизация производилась с помощью AdamW с learning rate 3e-4.

Обучение проводилось на GPU (CUDA), что позволило эффективно справляться с вычислительной нагрузкой. Модель оценивалась по метрикам accuracy, F1-score, ROC-AUC и PR-AUC, что обеспечило объективную оценку качества работы модели в условиях неравномерного распределения классов.

Полученные результаты показали высокую эффективность подхода на основе BERT в задаче определения токсичности текста и подтвердили его применимость в реальных сценариях автоматической модерации и фильтрации пользовательского контента.