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 1**

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**

Реализовать алгоритмы TF-IDF + word2vec (чтобы можно было складывать и вычитать слова). Также необходимо найти датасет для обучения написанных алгоритмов)

1. **Ход работы**

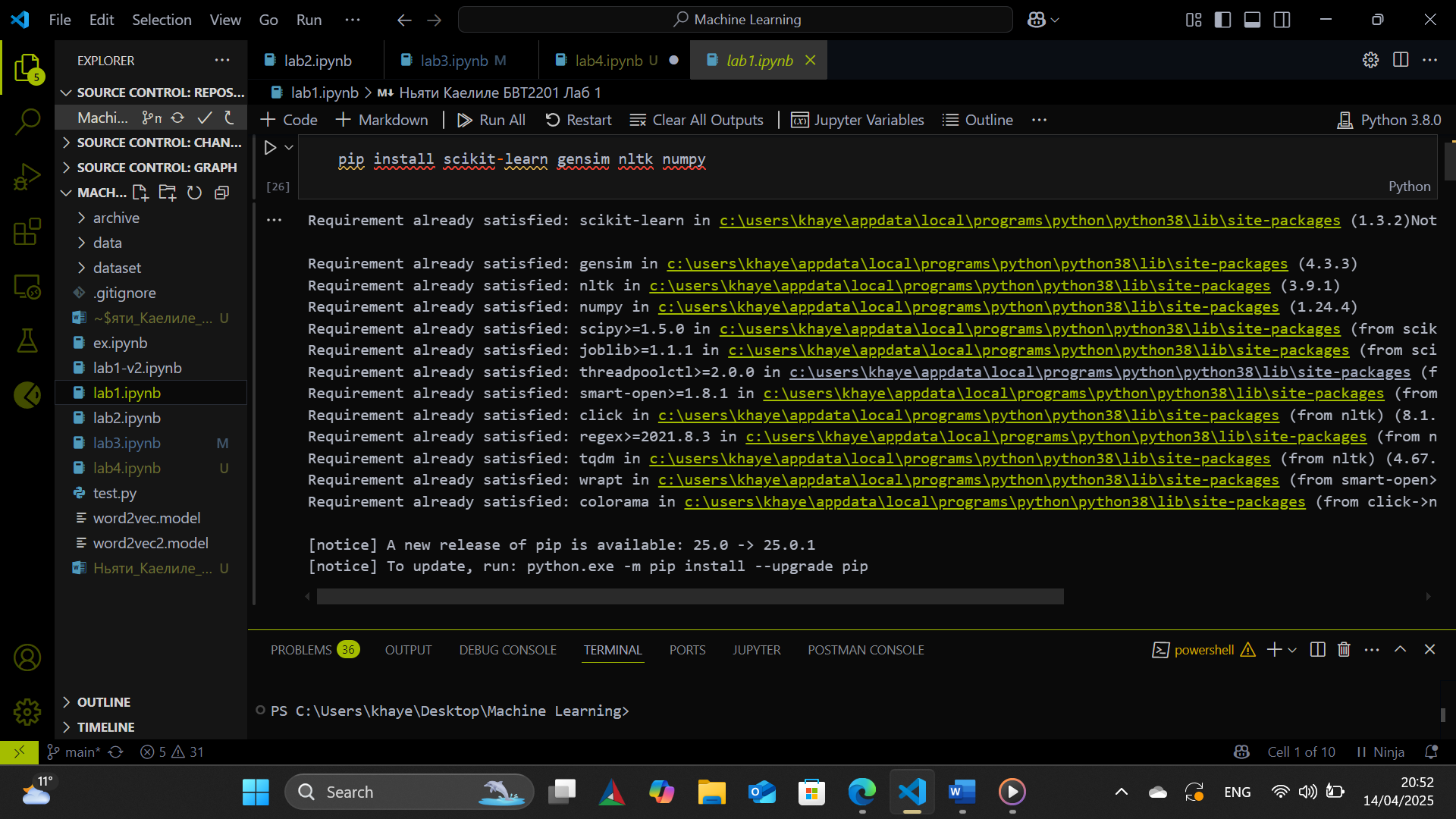
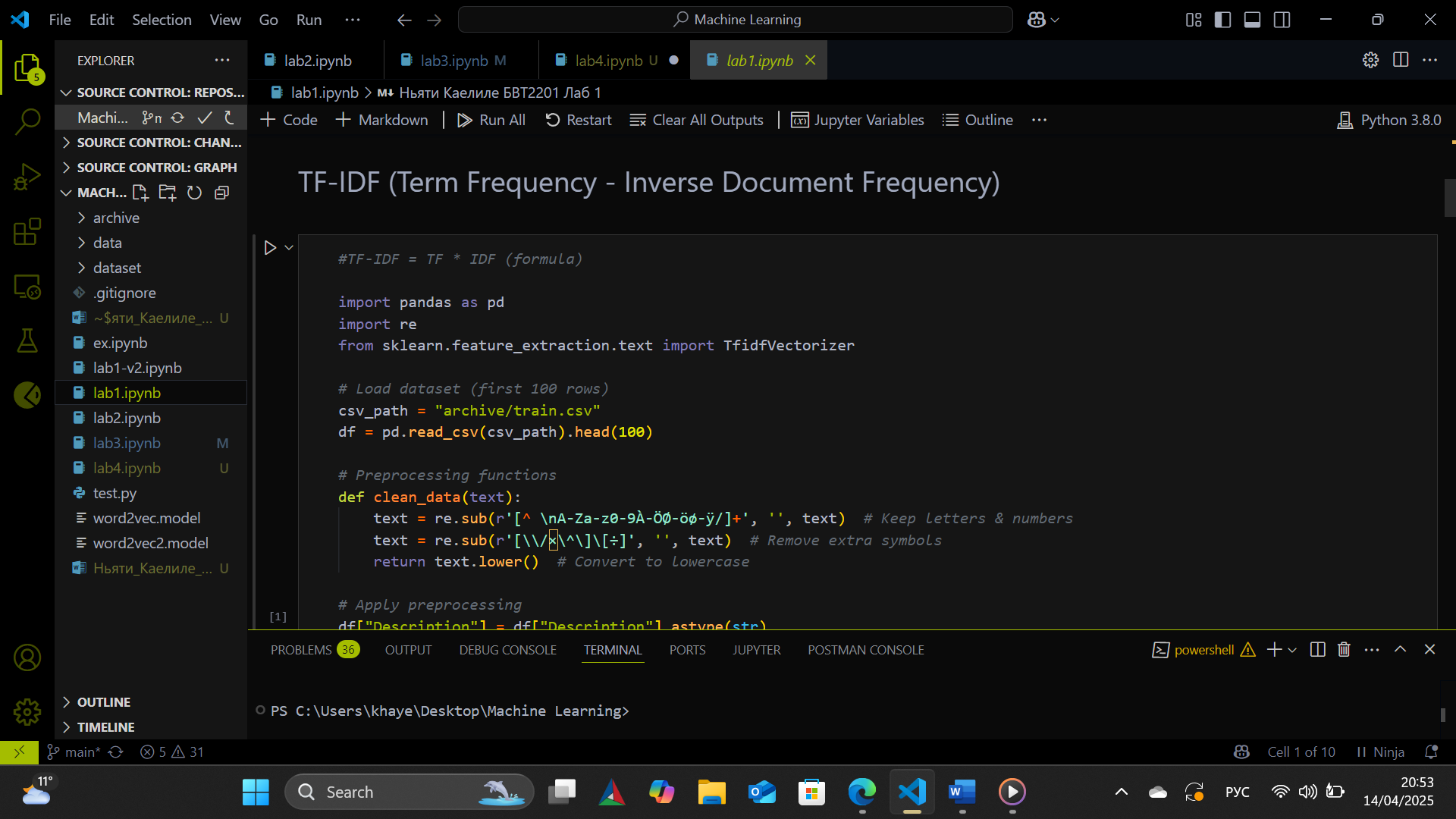
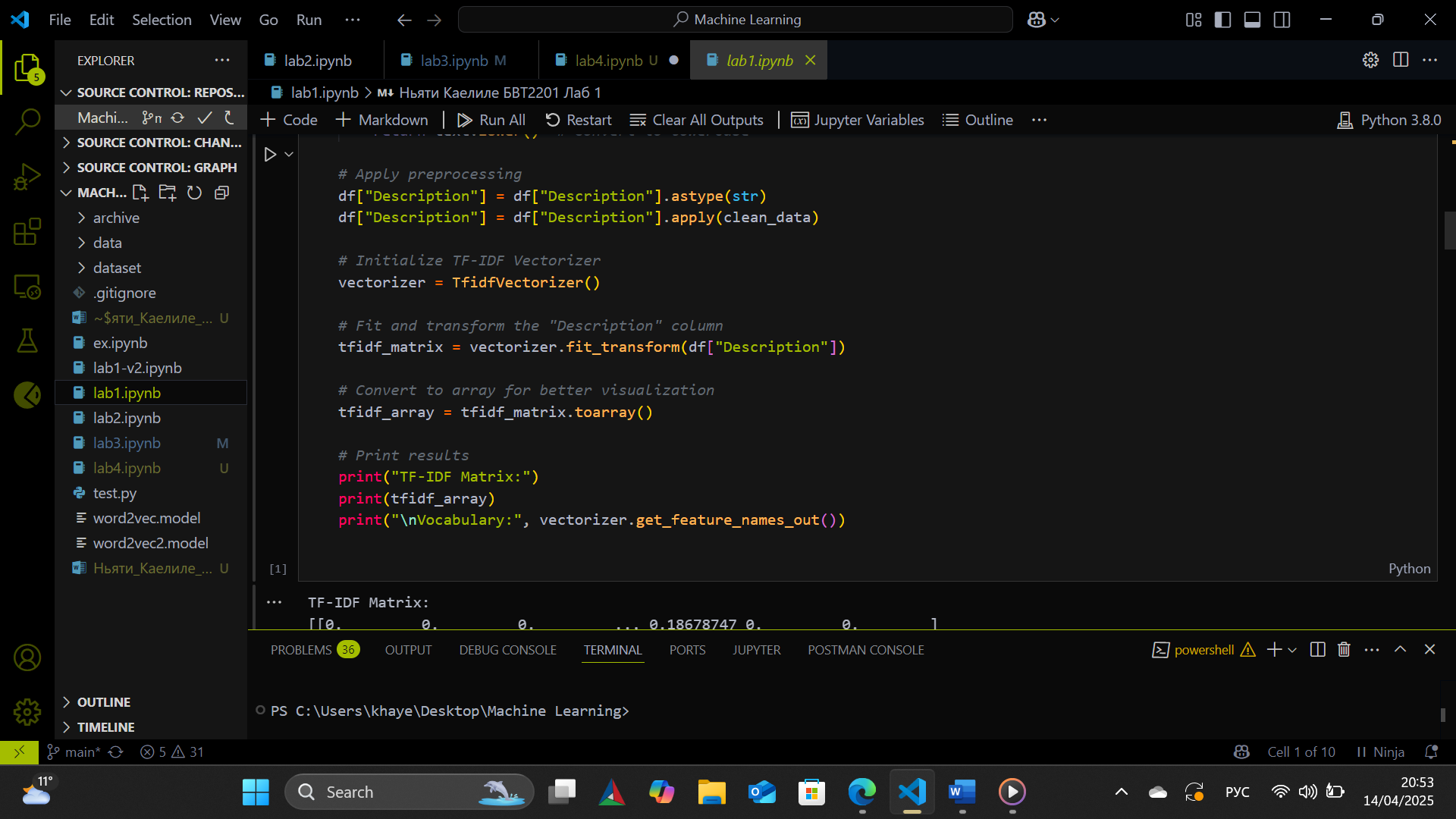


Рис 1. Библиотекы





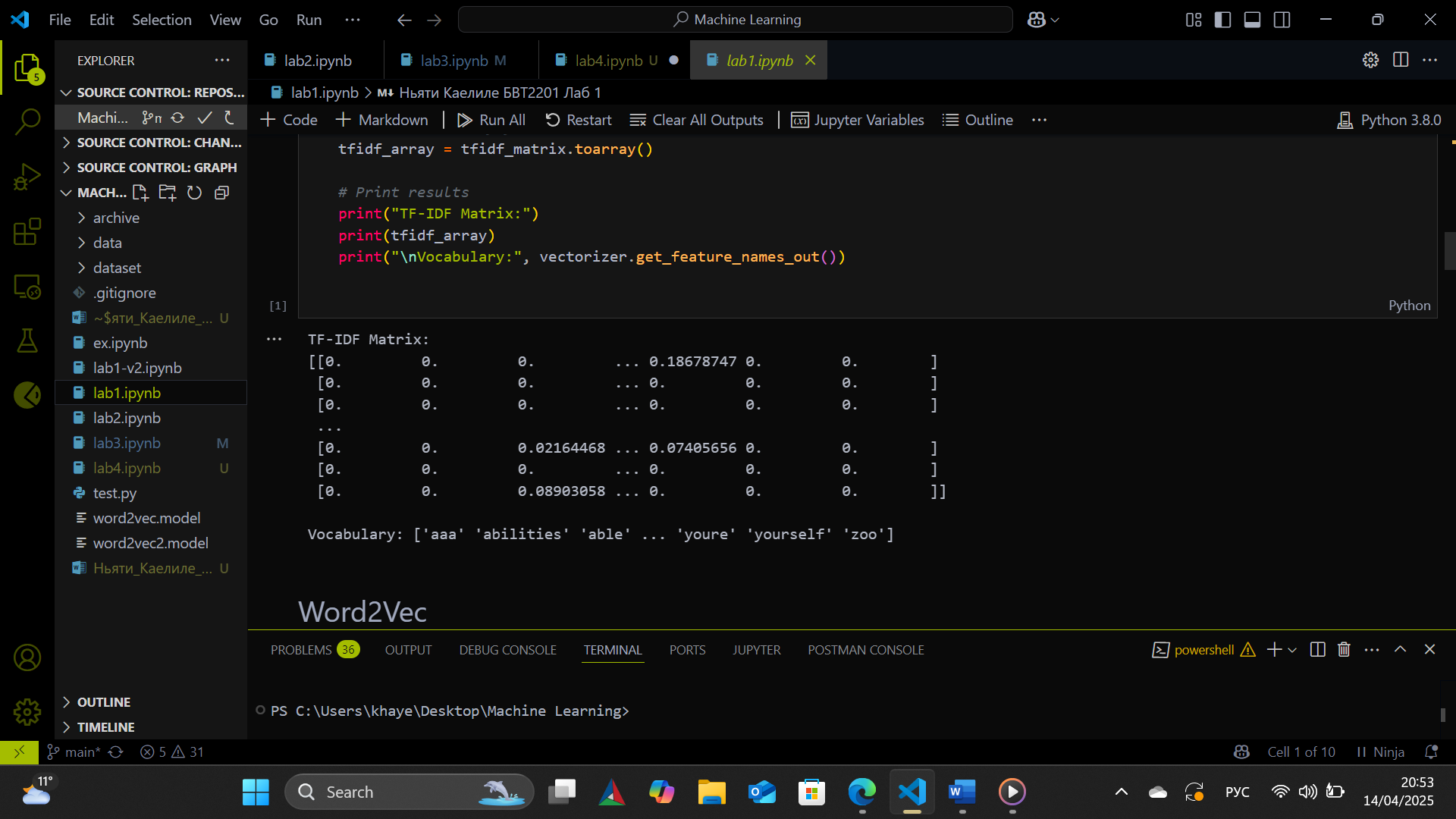
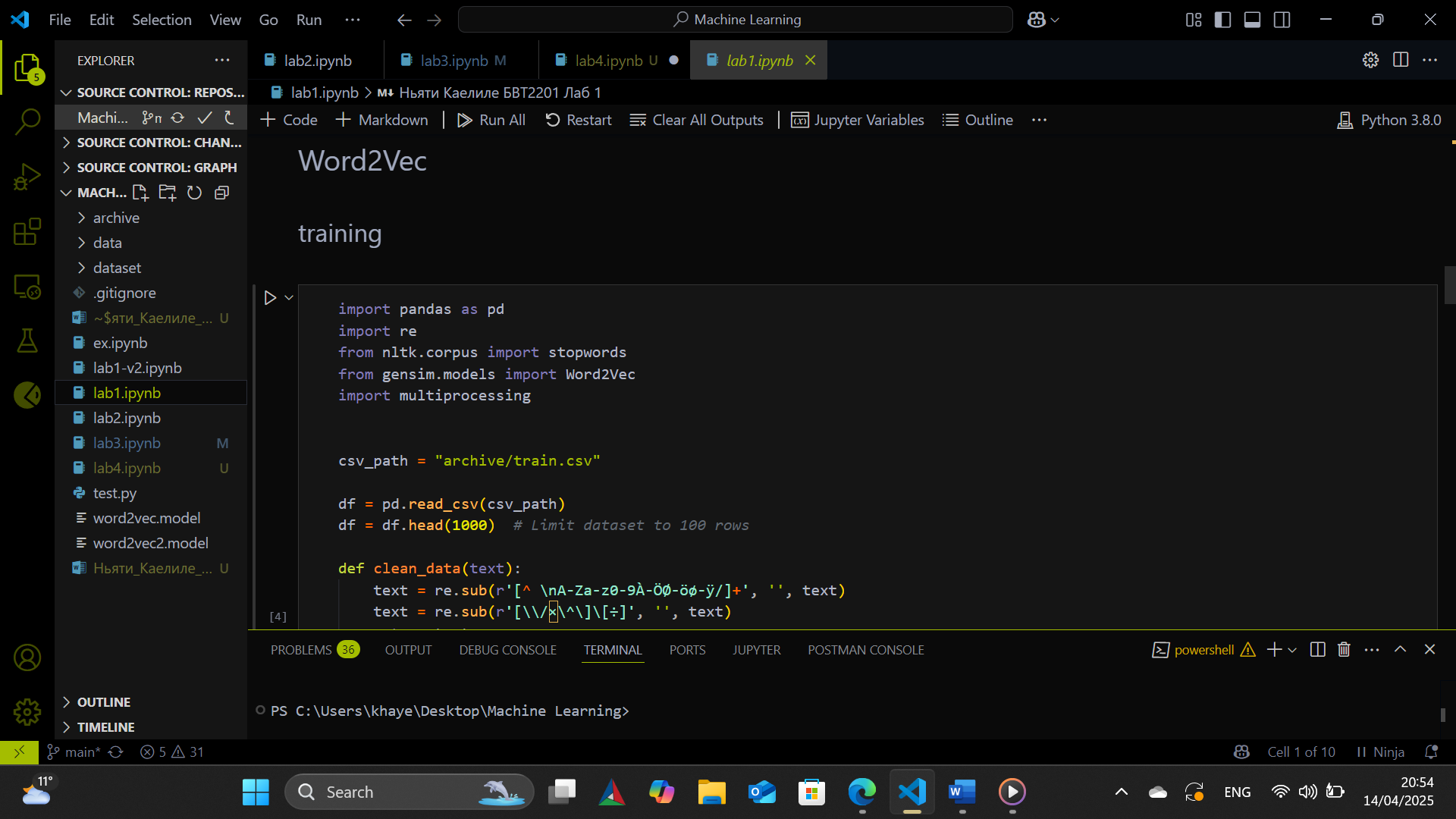
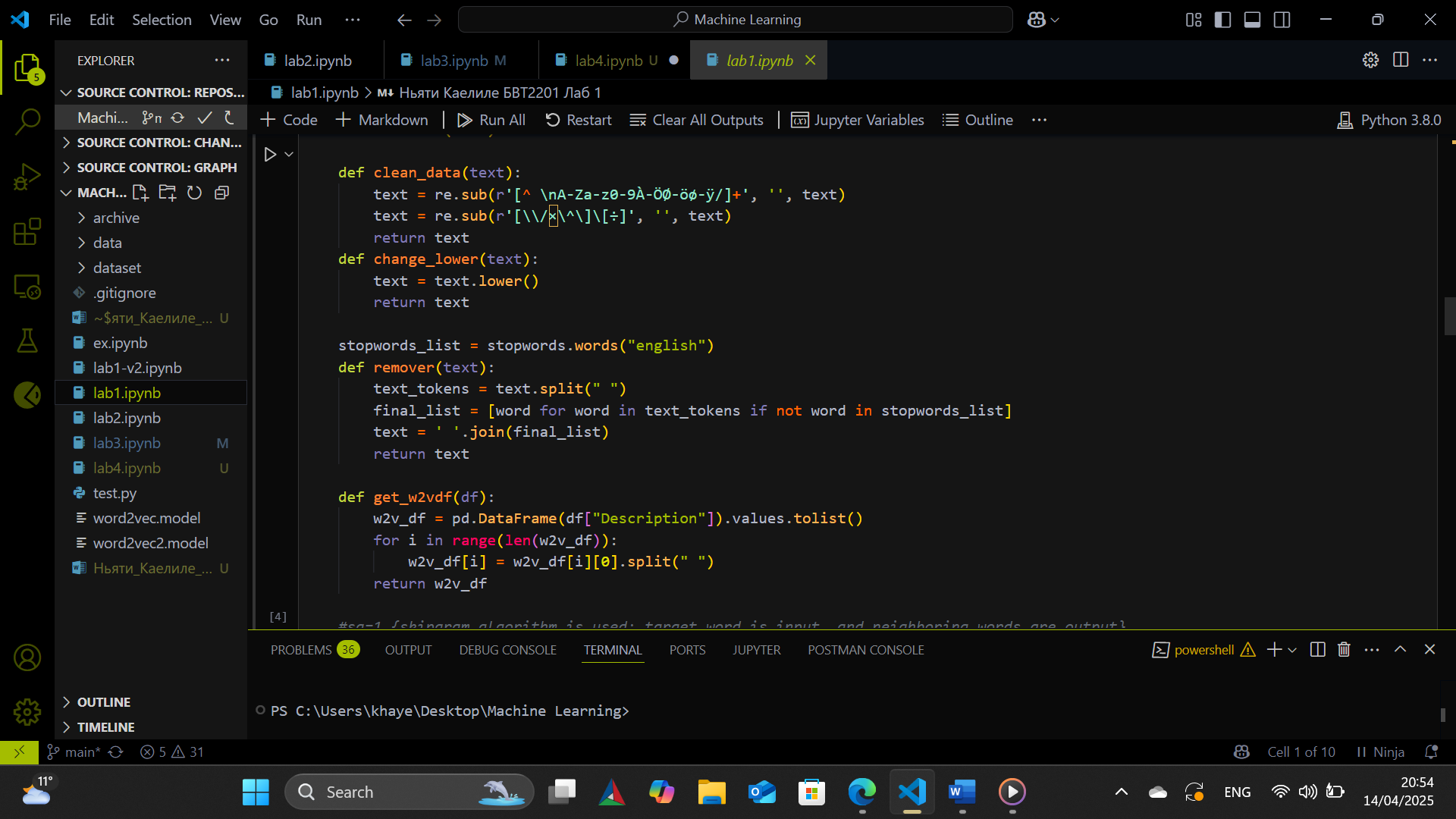
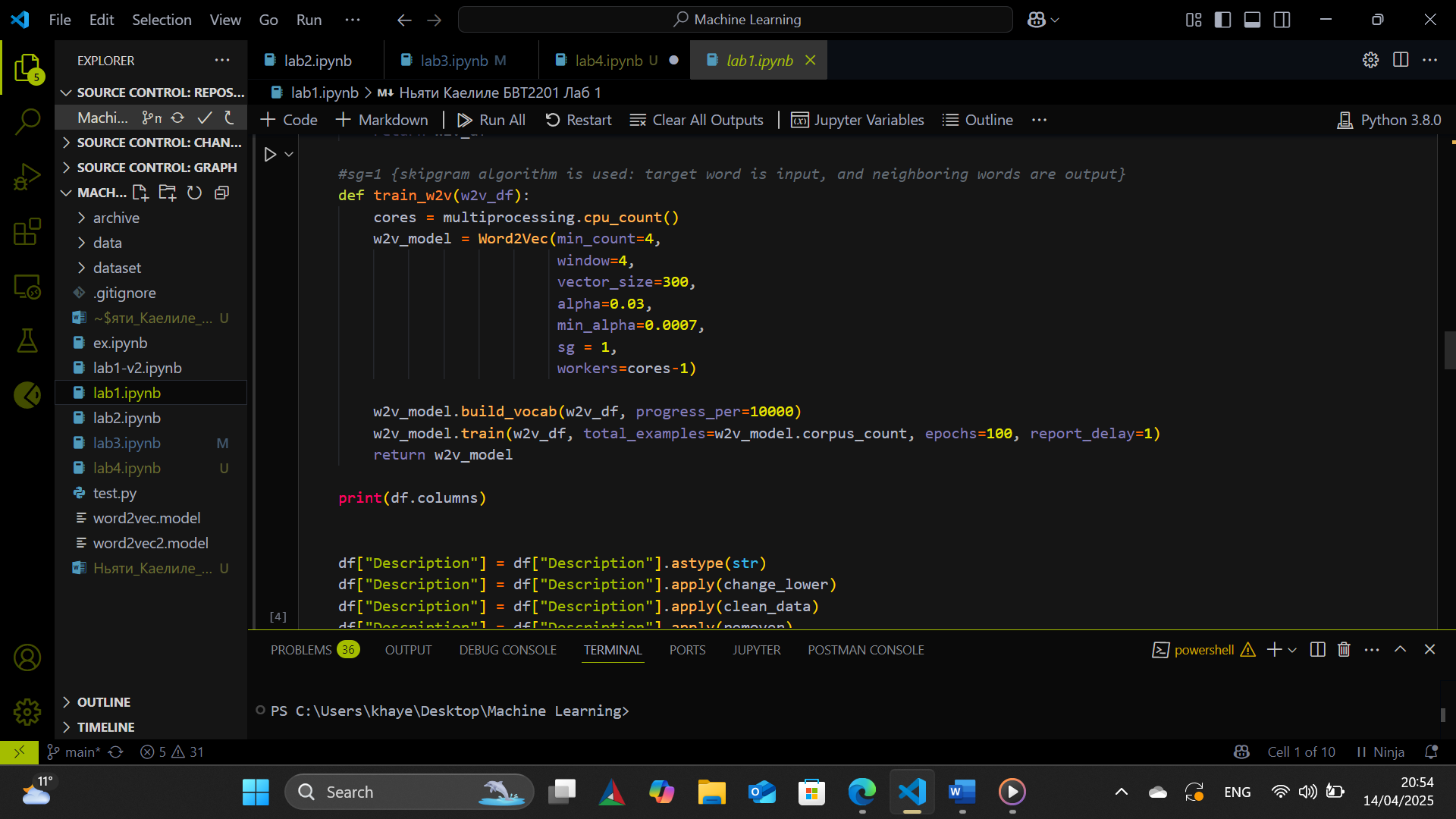


Рис 2. TF-IDF







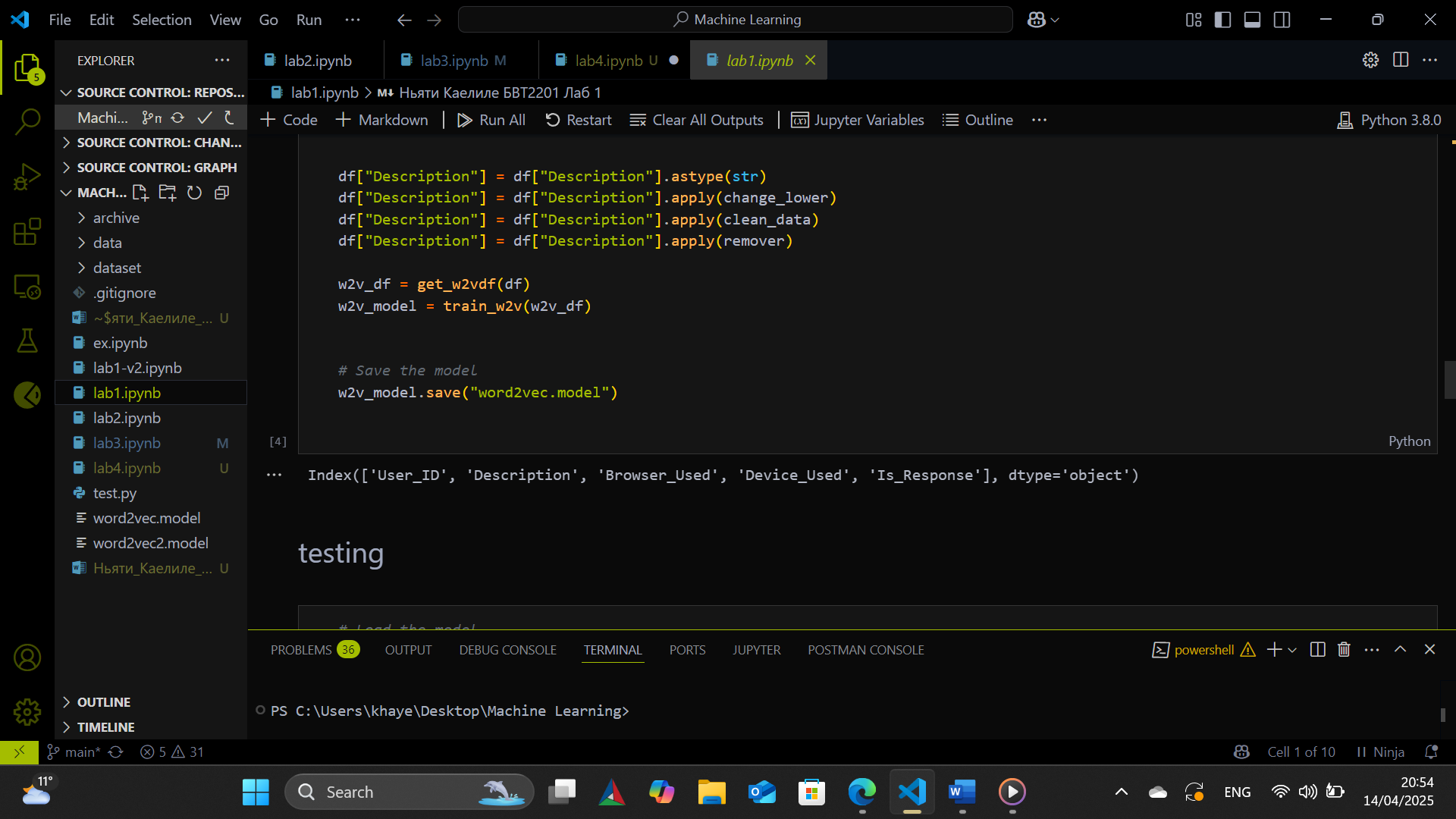


Рис 3. Word2Vec

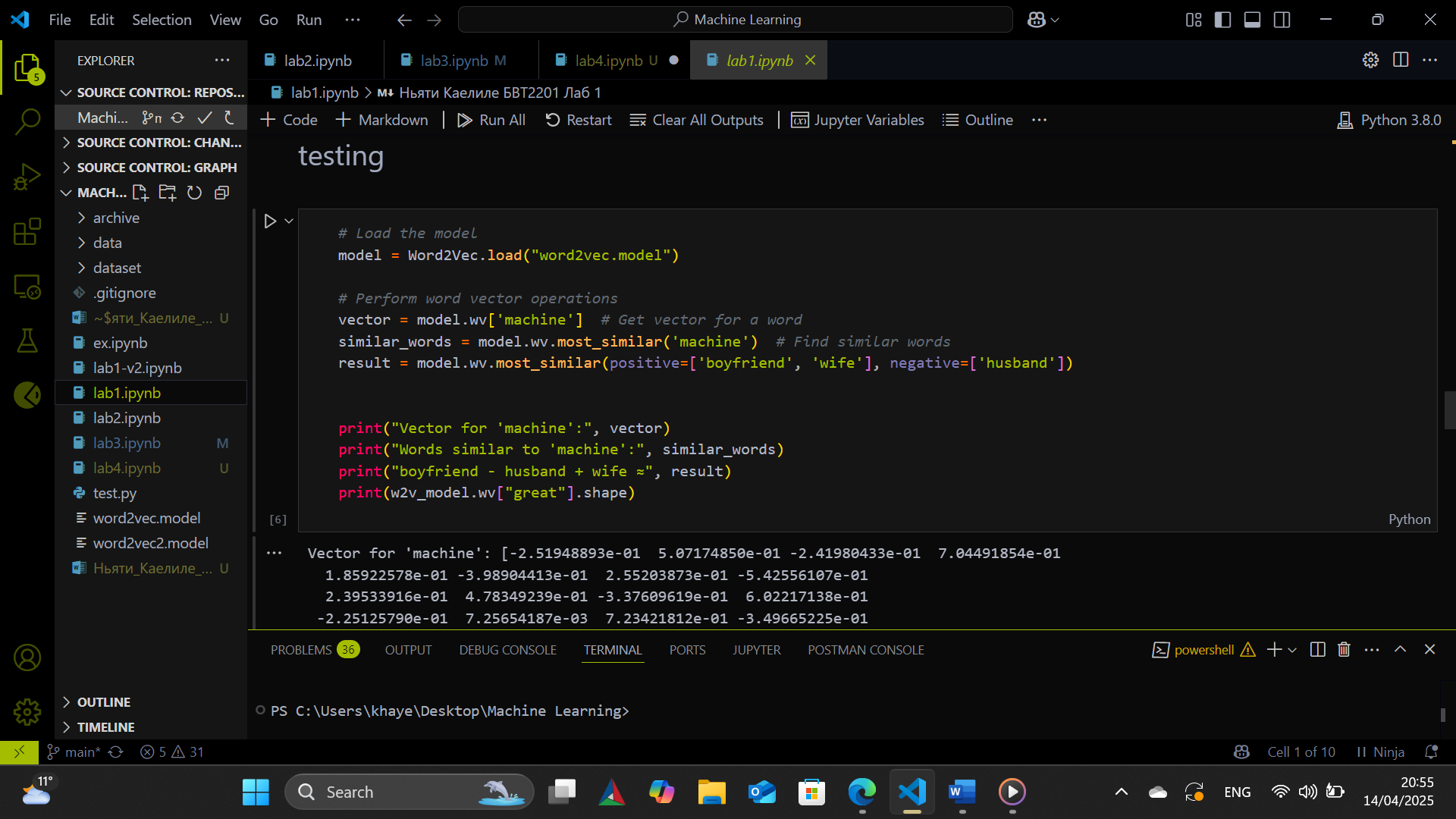
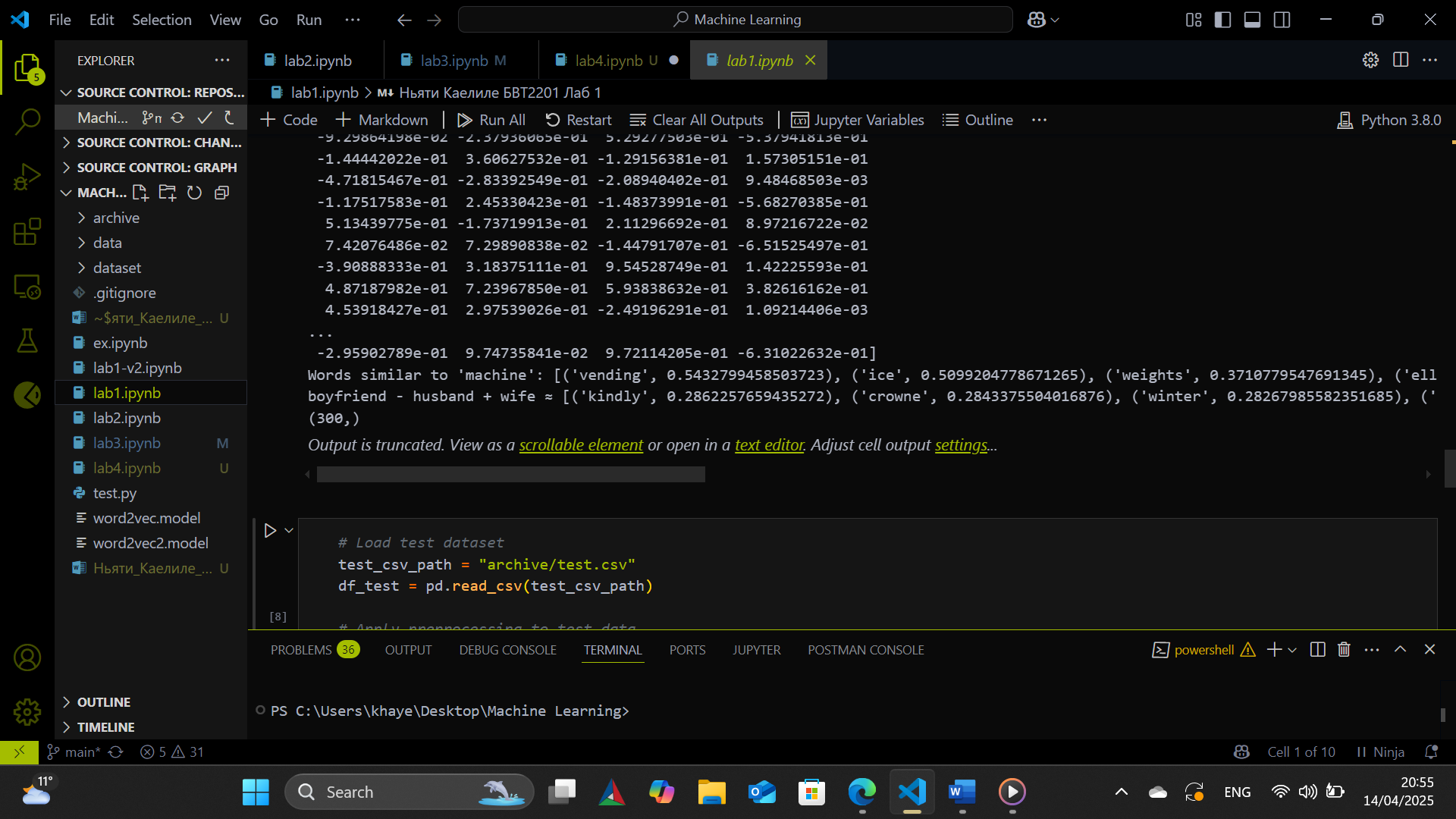
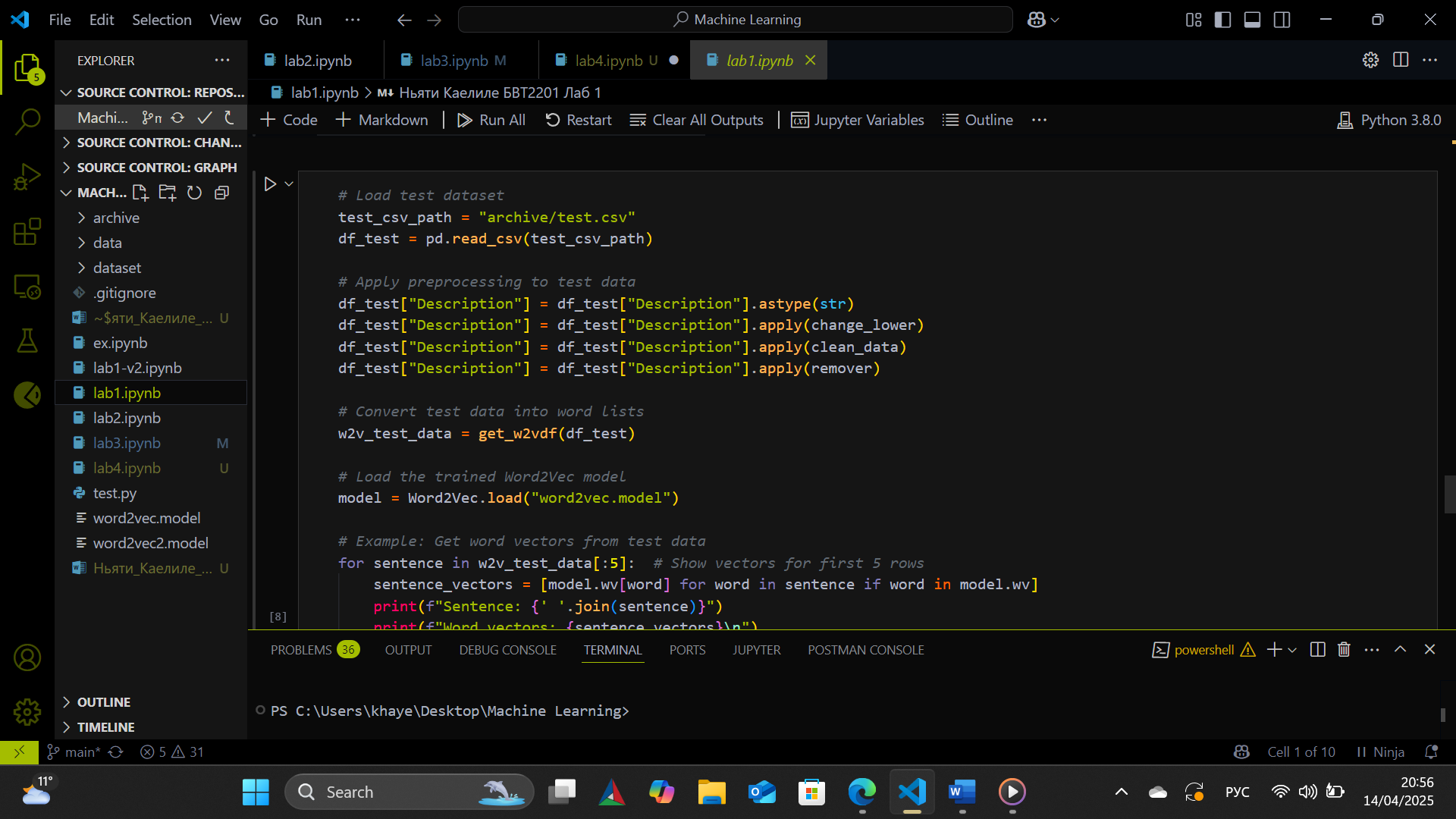
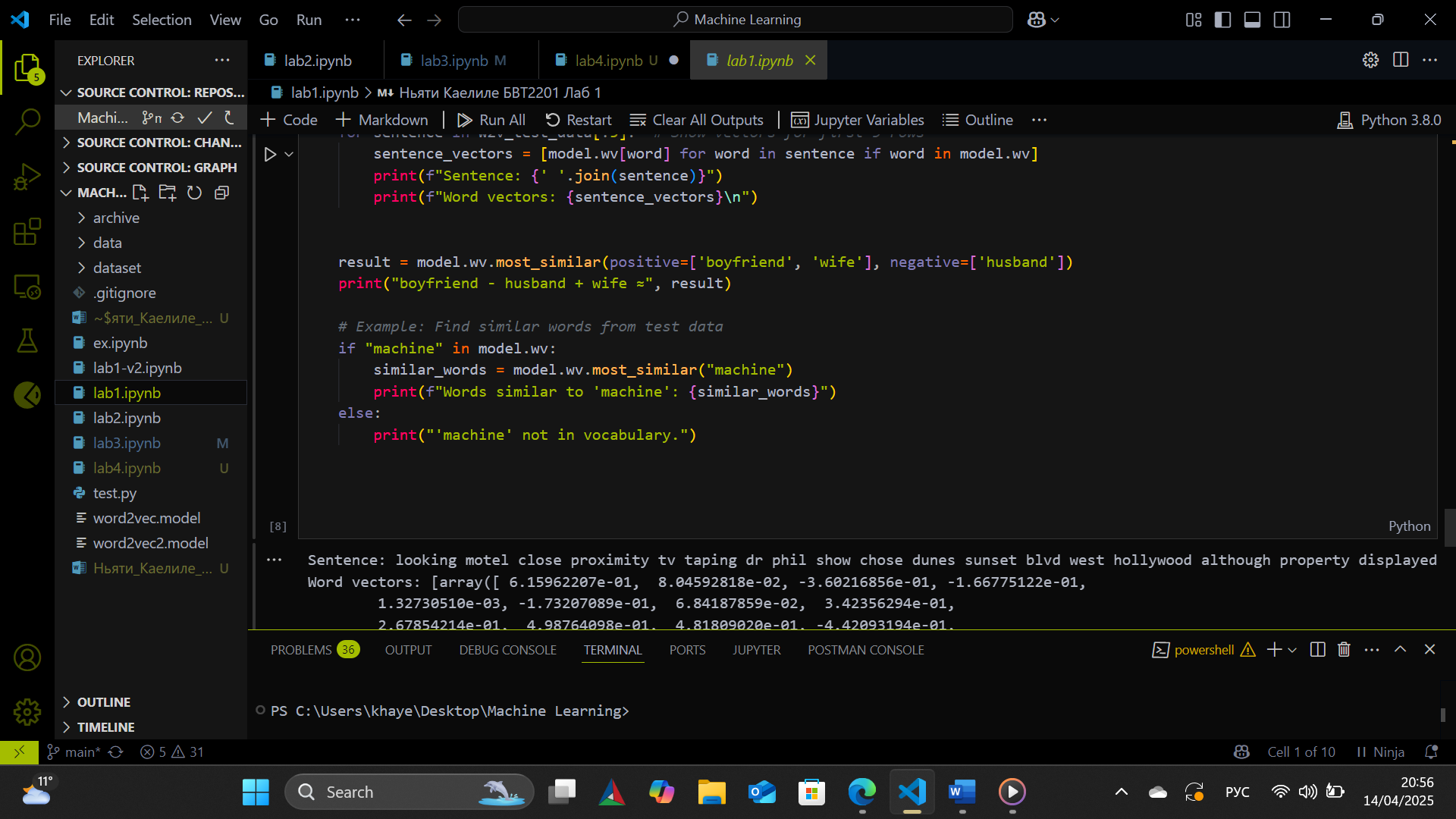
 

Рис 4. Тестирование





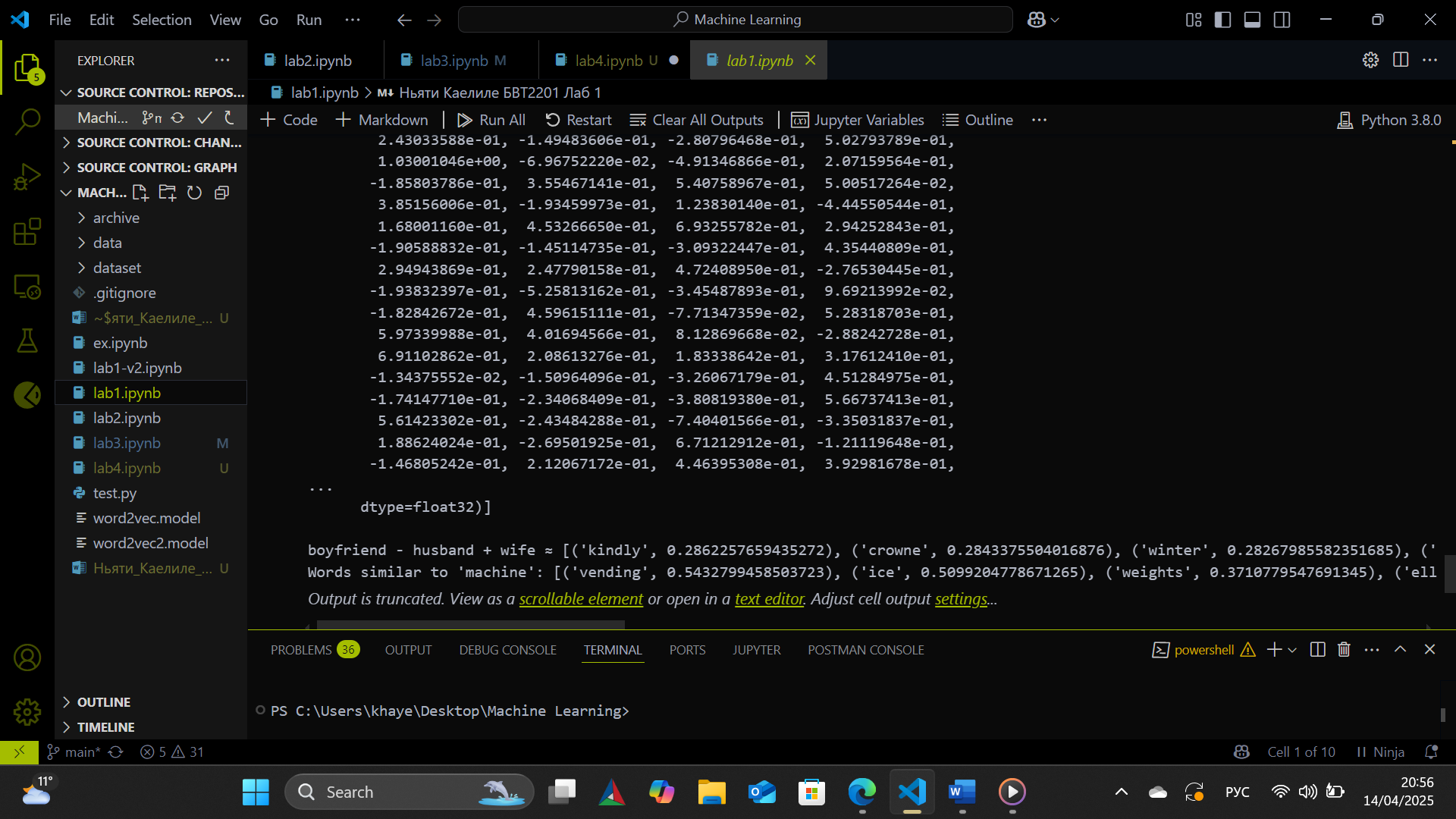


Рис 5. Результаты

1. **Выводы**

В рамках поставленной задачи была реализована комбинация алгоритмов TF-IDF и word2vec, позволяющая не только учитывать важность слов в контексте (TF-IDF), но и оперировать их семантическими представлениями (word2vec), что дает возможность производить операции над словами, такие как сложение и вычитание в векторном пространстве.

Для обучения и тестирования алгоритмов был подобран соответствующий датасет, обеспечивающий достаточное разнообразие лексики и контекстов. Такая интеграция подходов позволяет эффективно решать задачи семантического анализа и обработки естественного языка.