

**Технически университет - София**

Факултет по приложна математика и информатика

*София 2023г.*

Проверил:

Изработила:

Савина Вълчанова

Фак. Номер: 471220071

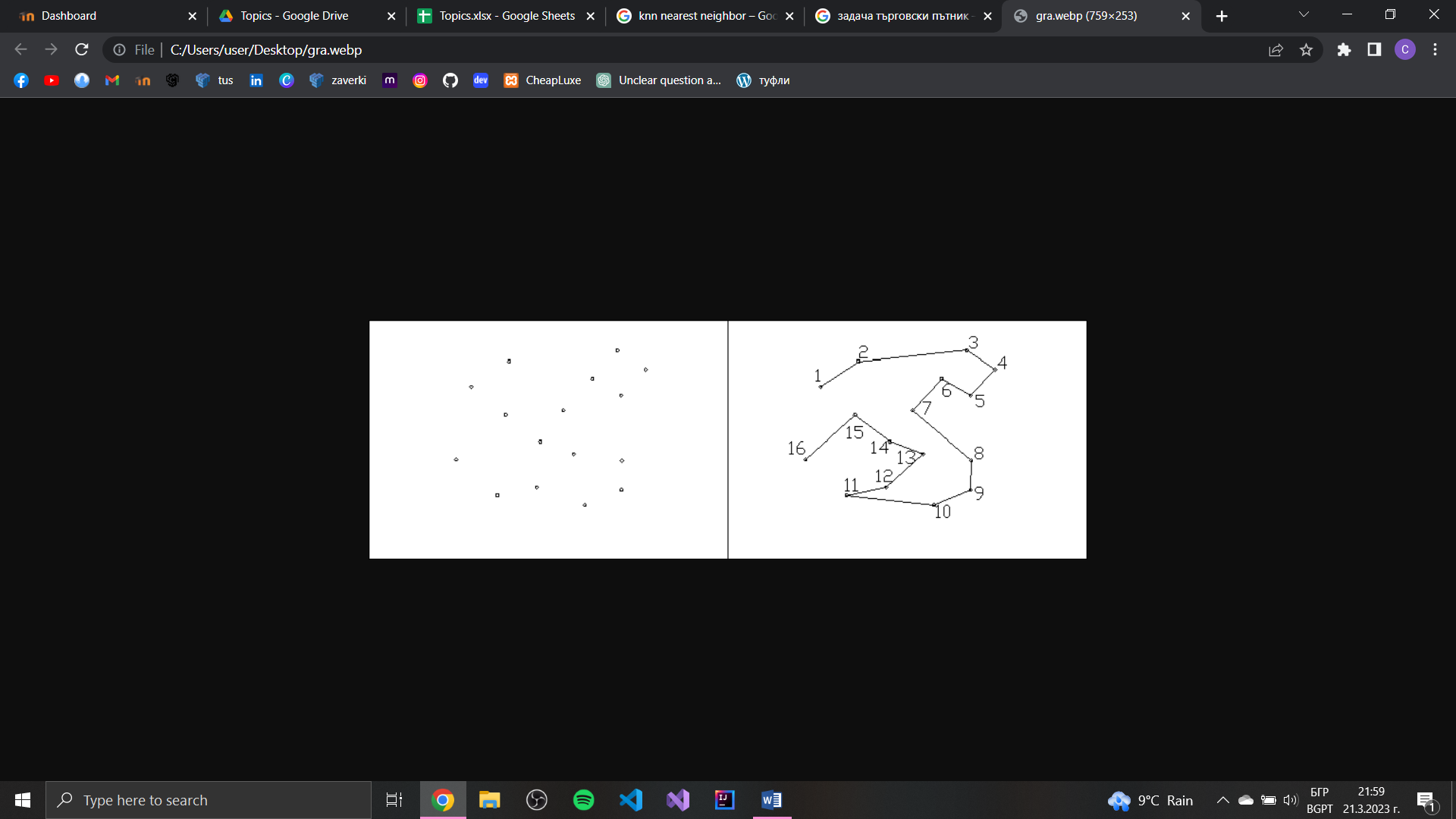
На тема:

K-Nearest Neighbors. Introduction, algorithm, modifications and applications

Курсова работа

1. **Теоретична част**
   1. Въведение

К най-близък съсед е популярен и прост алгоритъм, използван за проблеми с класификация и регресия. Базира се на идеята за намиране на K най-близки точки от данни в набора за обучение до нова точка от данни и след това предвиждане на етикета на тази нова точка от данни въз основа на етикетите на най-близките й съседи. KNN е непараметричен алгоритъм, което означава, че не прави никакви предположения относно основното разпределение на данните. Алгоритъмът открива кратък път (последователност от градове), но не задължително този път да е оптималният. Методът на най-близкия съсед е лесен за реализация и бързо се изпълнява, но поради характера си на алчен алгоритъм, понякога може да пропусне много по-кратки пътища в графа, които могат да бъдат забелязани интуитивно с човешко око.



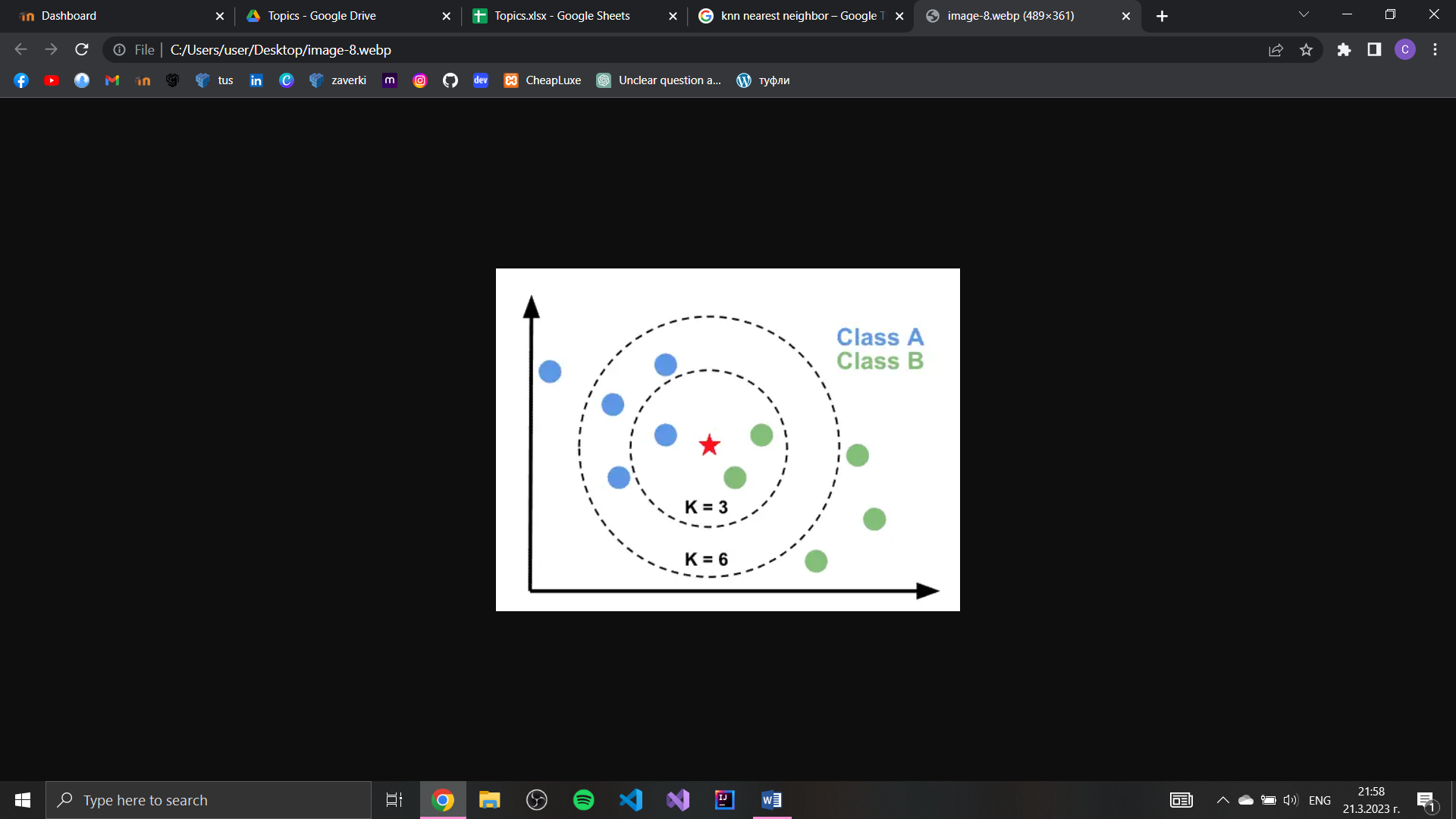
1.2. Алоритъм

Алгоритъмът на KNN може да бъде обобщен в следните стъпки:

* Изчислете разстоянието между новата точка от данни и всички точки от данни в набора за обучение.
* Изберете K най-близките точки от данни до новата точка от данни въз основа на изчислените разстояния.
* За проблеми с класификацията присвоете етикета на класа на новата точка от данни въз основа на мажоритарния клас на нейните K най-близки съседи. За проблеми с регресията присвоете изходната стойност на новата точка от данни въз основа на средната стойност на нейните K най-близки съседи.

Или:

* Избор на произволен възел за начален.
* Намиране на най-близкият непосетен възел
* Избор на възел V като текущ.
* Маркиране на възел V, като посетен.
* Ако всички възли са посетени, алгоритъма завършва.
* Връщане в стъпка 2.



1.3. Модификации

Има няколко модификации, които могат да бъдат направени в основния алгоритъм на KNN, за да се подобри неговата производителност:

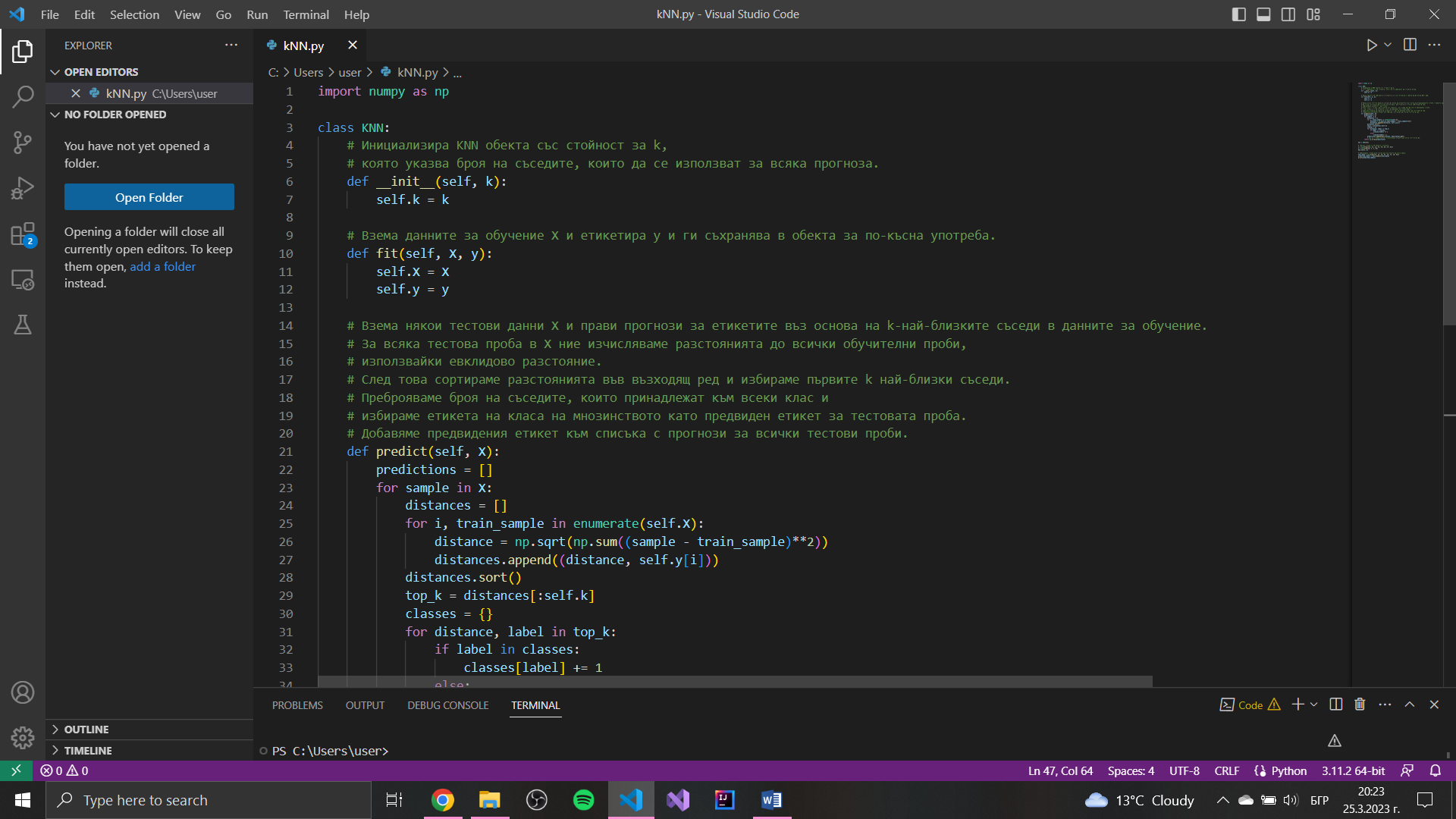
* Показател за разстояние: Изборът на показател за разстояние може да окаже значително влияние върху производителността на алгоритъма. Обикновено се използва евклидово разстояние, но могат да се използват и други показатели за разстояние, като разстоянието на Манхатън и косинусното сходство.
* Претеглен KNN: Вместо да присвоява равни тегла на всички K най-близки съседи, претегленият KNN присвоява по-високи тегла на по-близките съседи и по-ниски тегла на по-далечните съседи.
* Избор на функции: Избирането на най-подходящите функции може да подобри производителността на KNN чрез намаляване на размерността на данните и елиминиране на шума.
* Локализиран KNN: Традиционните алгоритми използват глобална стойност на k, която е еднаква за всички точки на заявка. Локализиранияt KNN обаче позволява стойността на k да варива в зависимост от локалната плътност на точките от данн около точката на заявката. Това може да подобри работата на KNN в области с висока или ниска плътност.
* Ансебли методи.

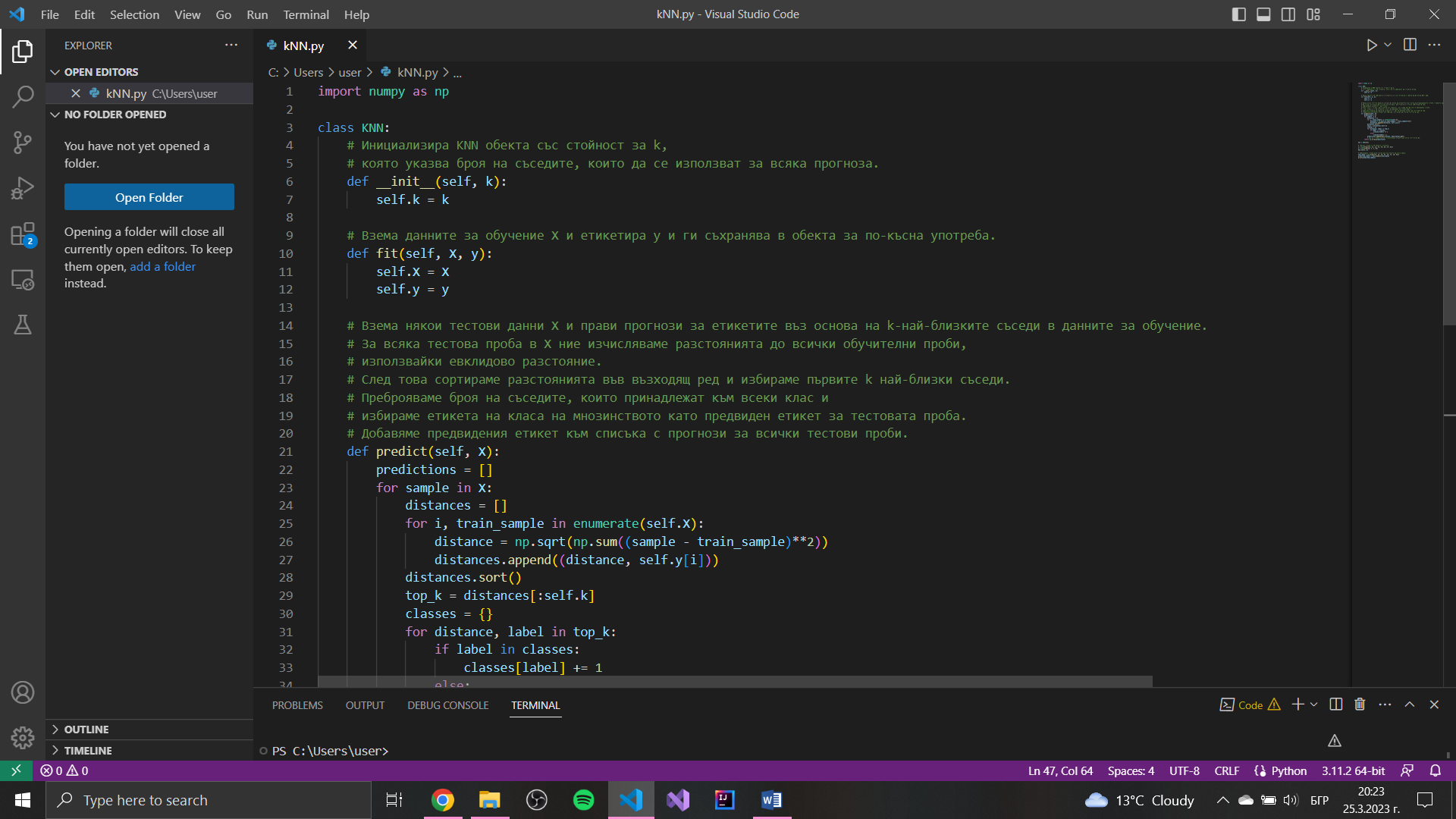
Изборът на модификации ще зависи изцяло от конкретния проблем и наличните данни.

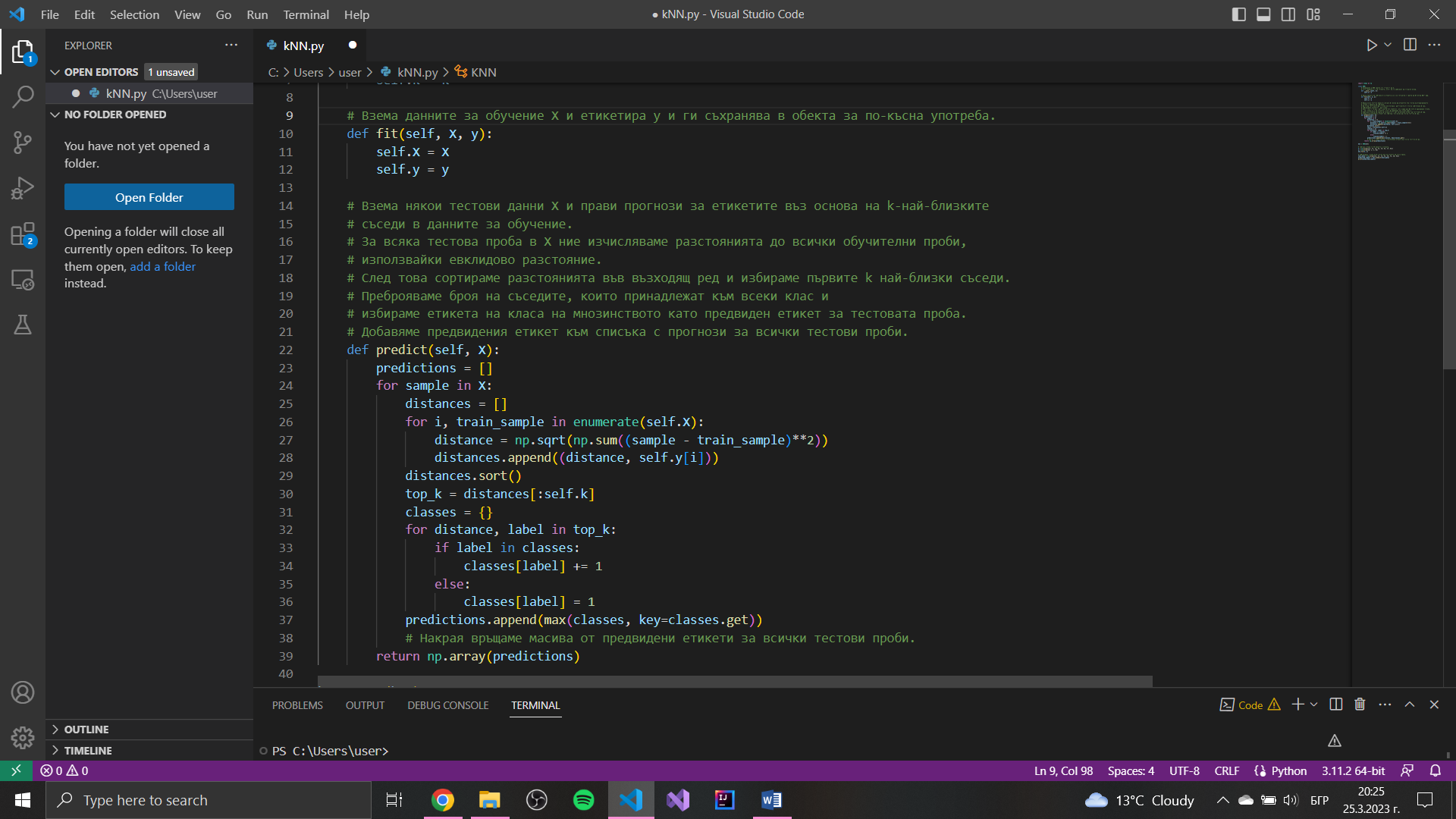
1.4. Приложения

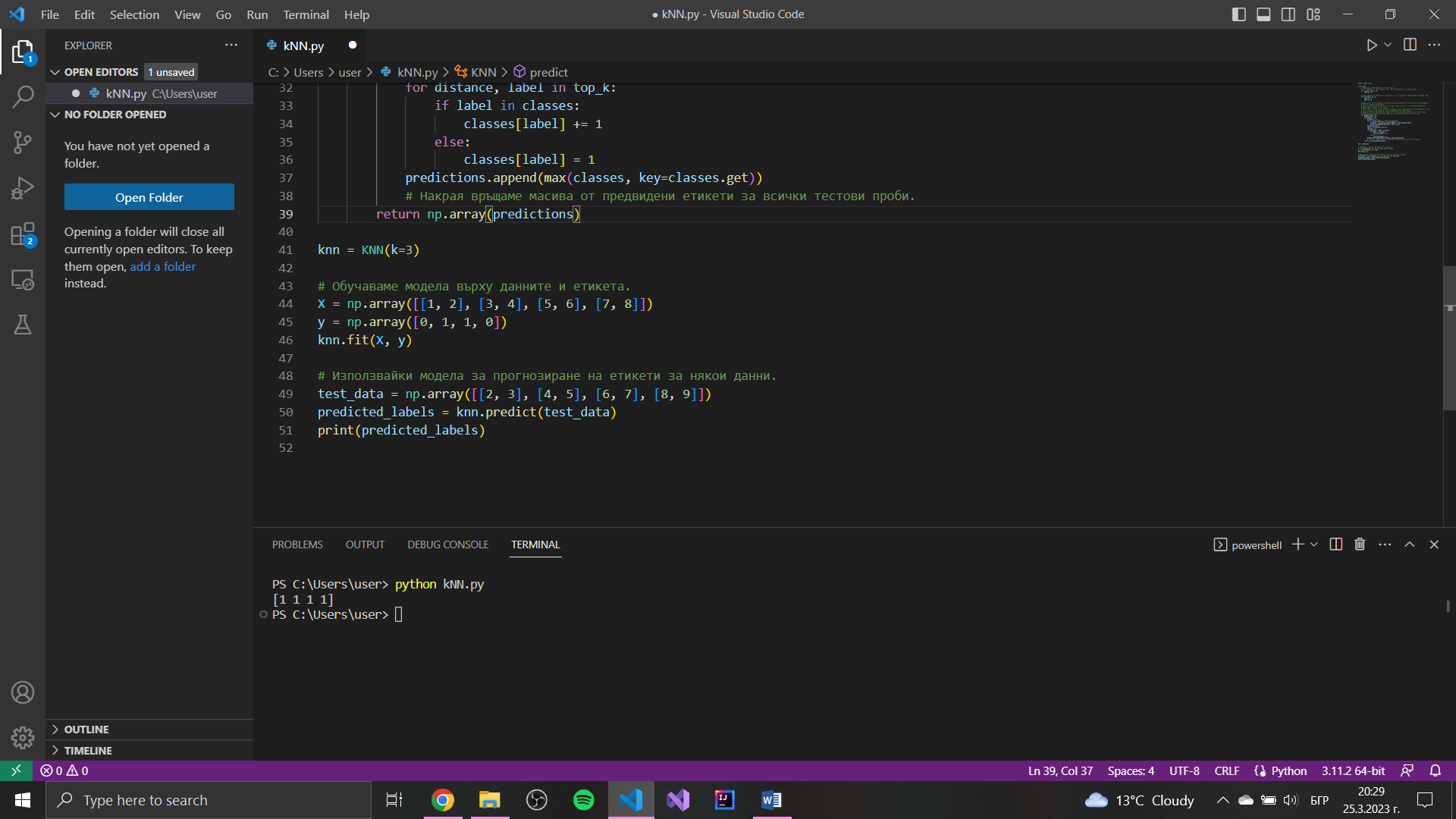
* Разпознаване на изображения: KNN може да се използва за класифициране на изображения въз основа на техните характеристики. Алгоритъмът може да се използва за класифициране на изображеия въз основа на техните характеристики, като цветове, форми и текстури.
* Системи за препоръчване: Изграждане на системи за препоръки, които предлагат продукти или услуги на потребителите въз основа на техните предпочитания и предишно поведение.
* Биоинформатика: KNN може да се използва за класифициране на ДНК последователности въз основа на техните характеристики.
* Откриване на измами: KNN може да се използва за откриване на измамни транзакции въз основа на моделите на минали транзакции.
* Класификация на текста: KNN може да се използва за задачи за класификация на текст като анализ на настроение, филтриране на спам и езикова идентификация. Алгоритъмът може да класифицира текст въз основа на сходство на думите и фразикте, използвани в текста.

1. **Имплементация**

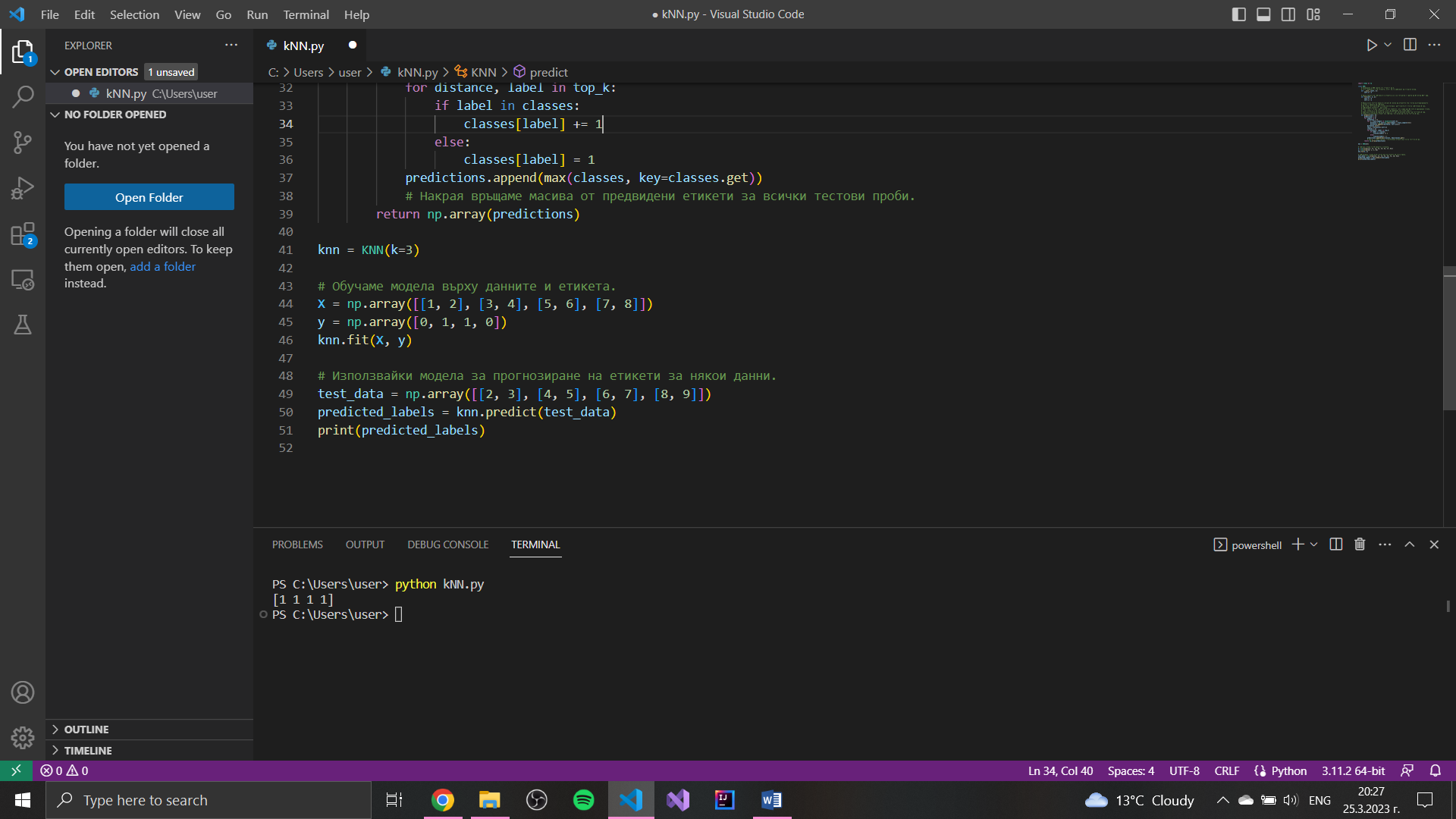








Резултат**:**



1. **Източници**

* <https://bg.wikipedia.org/wiki/%D0%9C%D0%B5%D1%82%D0%BE%D0%B4_%D0%BD%D0%B0_%D0%BD%D0%B0%D0%B9-%D0%B1%D0%BB%D0%B8%D0%B7%D0%BA%D0%B8%D1%8F_%D1%81%D1%8A%D1%81%D0%B5%D0%B4>
* <https://viktorgrigorov.wordpress.com/2017/02/26/290/>
* <https://chat.openai.com/?__cf_chl_tk=nbEBkUmXfmLxhZ79mYQ_b8nTU_UKiS.LAMBVTJQizU8-1679769042-0-gaNycGzNF2U>