КЛАСИФИКАЦИЈА СО Naïve Bayes

Целта на оваа лабораториска вежба е да се напише Баесов класификатор за класификација на студентите од втора година на ФИНКИ во категориите: “ќе дипломира на време”, “ќе дипломира со задоцнување”, “нема да дипломира”. Класификацијата треба да се направи врз основа на следните параметри: успех во средно образование (реален број во опсегот [2-5]), вид на средно образование (гимназиско или средно стручно), просек од прва година (реален број во опсегот [5-10]) и број на неположени предмети од прва година (0, од 1-2 предмети, од 3-5 предмети или повеќе од 5).

Тренинг множеството е составено од следните податоци:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Дипломирал** | **Успех (средно)** | **Вид на средно образование** | **Просек од прва година** | **Број на не положени предмети од прва година** |
| на време | 4.5 | гимназиско | 8.31 | од 1-2 |
| на време | 5 | гимназиско | 10 | 0 |
| на време | 5 | средно стручно | 7.98 | 0 |
| на време | 4.98 | гимназиско | 9.42 | 0 |
| на време | 4 | гимназиско | 8.98 | од 1-2 |
| со задоцнување | 4.23 | гимназиско | 8.23 | од 3-5 |
| со задоцнување | 3 | средно стручно | 8.56 | од 1-2 |
| со задоцнување | 5 | гимназиско | 10 | од 3-5 |
| со задоцнување | 4.87 | гимназиско | 7.99 | повеќе од 5 |
| со задоцнување | 4.56 | средно стручно | 8.98 | од 3-5 |
| не дипломирал | 3.92 | средно стручно | 7 | од 3-5 |
| не дипломирал | 3.45 | средно стручно | 5 | повеќе од 5 |
| не дипломирал | 2.67 | средно стручно | 6.3 | од 3-5 |
| не дипломирал | 3.2 | средно стручно | 8.2 | повеќе од 5 |
| не дипломирал | 2 | гимназиско | 6.2 | повеќе од 5 |

1. Да се пополни табелата за веројатностите за секоја од карактеристиките и категориите (да се земе во предвид и Лапласовото израмнување):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Успех (средно)** | | **Вид на средно обр.** | |
| **mean** | **variance** | **гимназиско** | **средно стручно** |
| **на време** | 4.696 | 0.397 | 0.8 | 0.2 |
| **со задоцнување** | 4.332 | 0.717 | 0.6 | 0.4 |
| **не дипломирал** | 3.048 | 0.707 | 0.2 | 0.8 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Просек од прва година** | | **Број на неположени предмети од прва година** | | | |
| **mean** | **variance** | **0** | **од 1-2** | **од 3-5** | **повеќе од 5** |
| **на време** | 8.938 | 0.731 | 0.542 | 0.375 | 0.042 | 0.042 |
| **со задоцнување** | 8.752 | 0.707 | 0.042 | 0.208 | 0.542 | 0.208 |
| **не дипломирал** | 3.048 | 0.661 | 0.042 | 0.042 | 0.375 | 0.542 |

1. Да се напише програма за класификаторот (со кратки коментари) во јазик по ваша желба (matlab, octave, python, java, дури и excel…)

1. Да се напише излез од програмата при внесување на следните тест примероци

|  |  |  |  |
| --- | --- | --- | --- |
| **Успех (средно)** | **Вид на средно образование** | **Просек од прва година** | **Број на не положени предмети од прва година** |
| 4.7 | гимназиско | 9 | од 1-2 |
| 3 | гимназиско | 9 | од 1-2 |
| 3 | средно стручно | 8 | повеќе од 5 |

0: 83.538%

1: 16.459%

2: 0.003%

0: 0.266%

1: 97.891%

2: 1.843%

0: 0.001%

1: 6.546%

2: 93.453%