**Lineárna regresia a umelá neurónová sieť**

**Úloha:**

Porovnať výkonnosť klasifikačných modelov, konkrétne logistickej regresie a umelej neurónovej siete, na dvoch rôznych datasetoch - "Social\_Network\_Ads.csv" a "Churn\_Modelling.csv".

**Datasety:**

**Social\_Network\_Ads.csv**: Obsahuje informácie o používateľoch, ako vek a odhadovaný príjem, a cieľovú premennú, ktorá určuje, či používateľ zakúpil produkt (1) alebo nie (0).

**Churn\_Modelling.csv**: Obsahuje údaje o zákazníkoch banky, vrátane demografických informácií, úrovne kreditu, počtu produktov, a závislej premennej, ktorá určuje, či zákazník opustil banku (1) alebo zostal (0).

**Postup:**

**Logistická regresia:** Logistická regresia bola použitá pre oba datasety, pričom výsledky boli vyhodnotené pomocou confusion matrix a accuracy skóre.

**Umelá neurónová sieť (ANN):**

ANN bola implementovaná iba pre oba datasety. Sieť obsahovala dve skryté vrstvy s 6 neurónmi, relu aktivačnou funkciou a jednu výstupnú vrstvu s sigmoid funkciou.

**Výsledky:**

1. **Social\_Network\_Ads.csv:**

**Logistická regresia:**

A diagram of a training set

Description automatically generatedA green and red graph

Description automatically generatedLogistická regresia dosiahla určitú presnosť pri klasifikácii, pričom jej výsledky boli vizualizované na základe dvoch nezávislých premenných (vek a odhadovaný príjem). Vizualizácia rozhodovacej hranice ukazuje, ako dobre model oddeľuje dve triedy.

**Confusion matrix**

|  |  |
| --- | --- |
| **65** | **3** |
| **8** | **24** |

**Accuracy = 0.89**

A graph of a graph

Description automatically generated**ANN:**Trénovanie neurónovej siete zahrňalo graf presnosti modelu počas jednotlivých epoch. Výsledná confusion matrix ukázala výrazne lepšiu presnosť oproti logistickej regresii, čo poukazuje na schopnosť neurónovej siete lepšie sa naučiť zložitejšie vzorce.

**Confusion matrix**

|  |  |
| --- | --- |
| **64** | **4** |
| **3** | **29** |

**Accuracy = 0.93**

1. **Churn\_Modelling.csv:**

**Logistická regresia:**

Aj v tomto prípade sme požili confusion matrix na hodnotenie výsledkov modelu. Logistická regresia nebola dostatočne flexibilná na zachytenie komplexných vzťahov medzi premennými, čo sa prejavilo mierne nižšou presnosťou.

**Confusion matrix**

|  |  |
| --- | --- |
| **1901** | **90** |
| **393** | **116** |

**Accuracy = 0.8068**

**ANN:**Tréning neurónovej siete ukázal postupný nárast presnosti s pridanými epochami. confusion matrix pre neurónovú sieť odhalila lepšie výsledky v porovnaní s logistickou regresiou, keďže model dokázal účinne identifikovať zložitejšie vzory.

A graph of a graph

Description automatically generated

**Confusion matrix**

|  |  |
| --- | --- |
| **1509** | **86** |
| **200** | **205** |

**Accuracy = 0.857**

**Diskusia:**

* V datasete “Social\_Network\_Ads.csv“ sú nezávislé premenné vek a príjem, kým závislá premenná je rozhodnutie o kliknutí na reklamu.
* V datasete “Churn\_Modelling.csv“ sú nezávislé premenné údaje o zákazníkoch ako vek, počet produktov a geografická poloha, kým závislá premenná predstavuje odchod zákazníka.

**Záver:**

Na základe porovnania sme zistili, že neurónové siete majú výraznú výhodu oproti logistickej regresii pri riešení zložitejších problémov, pretože lepšie zachytávajú nelineárne vzory v údajoch. Logistická regresia však môže byť vhodným riešením pre jednoduchšie problémy.