**Projektna naloga**

Pri predmetu MOS

**Avtor:** Urban Vižintin

**Smer študija:** ITK UNI

**Študijsko leto:** 2022/23

Bayesova klasifikacija

Bayesova klasifikacija je tip klasifikacije, ki temelji na Bayesovem zakonu verjetnosti. Bayesov zakon verjetnosti napove razred posamezne instance na podlagi že znane značilnosti podatkov. Tako izračuna verjetnost za vsak posamezen razred, ki je v podatkovni množici in se odloči za razred z največjo verjetnostjo, da najbolj ustreza instanci.

Bayesovo pravilo (verjetnost razreda C pri pogoju ):

“Naivno” Bayesovo pravilo (predpostavka neodvisnosti ):

Bayesov klasifikator je zelo uporabljen klasifikacijski algoritem. Prednosti Bayes-a so:

* Enostavna implementacija
* Učinkovit v visokodimenzijskih podatkovnih množicah (veliko parametrov)
* Enostavno dodajanje novih podatkov ter posodabljanje modela
* Možnost obvladovanja ter obravnave manjkajočih podatkov
* ...

Seveda pa ni perfekten algoritem. Ima tudi par pomankljivosti:

* Naivni Bayes se nanaša na to, da so spremenljivke neodvisne med seboj
* Občutljiv na nepomembne lastnosti
* Občutljiv na točnost podatkov (pravilnost)
* Za točne podatke potrebuje veliko količino podatkov
* ...

Podatkovna zbirka

Za podatkovno zbirko sem si izbral podatkovno množico iz spletne strani kaggle.com. Podatkovna zbirka se imenuje NASA – Nearest Earth Objects in vsebuje podatke o vesoljskih objektih, ki potujejo po vesolju. Zbrani so tudi podatki o teh objektih (ime, diameter, hitrost ...) ter napovedni atribut je vrednost, ali je ta vesoljski objekt nevaren za Zemljo (če lahko pride tako blizu Zemljini orbiti, da jo zadane oziroma pride tako blizu, da lahko povzroči kakšne efekte na Zemlji).

Podatkovna zbirka ima naslednje parametre:

* id -> edinstven identifikator za vsak posamezen asteroid / objekt
* name -> ime objekta, ki mu ga je dodelila NASA
* est\_diameter\_min -> minimalno ocenjen diameter v kilometrih
* est\_diameter\_max -> maksimalno ocenjen diamtere v kilometrih
* relative\_velocity -> hitrost glede na Zemljo
* miss\_distance -> zgrešena razdalja v kilometrih
* orbiting\_body -> planet, ki ga objekt obkroža (unikatna vrednost Zemlja)
* sentry\_object -> avtomatiziran sistem za spremljanje trkov (unikatna vrednost False)
* absolute\_magnitude -> opisuje notralno svetilnost
* hazardous -> napovedna vrednost
  + true -> objekt je nevaren
  + false -> objekt ni nevaren

Slika, ki vsebuje besede miza

Opis je samodejno ustvarjen

Namen podatkovne zbirke je na podlagi zbranih podatkov izdelati model, ki bi lahko v prihodnosti napovedal, če lahko bližajoči se asteroid vpliva na življenje na Zemlji in koliko je nevaren za nas.

Predstavitev rezultatov