Vlerësimi Makinave

Xhafer Muja

Fakulteti i Shkencave Kompjuterike, Universiteti “Ukshin Hoti” Prizren

Prizren, Republika e Kosovës

210306076.b@uni-prizren.com

**Abstrakti**

**Definimi i problemit:**  Ky problem daton që në vitin 1997 ,kur në atë kohë njerzit nuk mund të përcaktonin se a është e vlefshme një makinë për tu blerë , a i permbushte kërkesat specifike të klientëve dhe në baze te atyre detajeve vlerësohet nëse makina është e pranueshme për ato specifika te dhëna .

**Objektivat**: Për këtë problem ekzistojnë punime të shumta të personave te ndryshem ,dhe me perdorimin e alogirtmeve te ndryshme si . Random Forest , Naïve Bayess , Decision Trees , K-Nearest Neighbours(KNN).

**Metodologjia**: Metodologjia per këtë projekt ështe e ngjashme me metodologjitë e tjera , ne fillim kam pastruar datasetin , pastaj ne bazë të atyre vlerave me algoritmin e gatshëm Naïve Bayess trajnova modelin dhe ne fund nxora rezultatet.

**Rezultatet e pritshme**: Pas trajnimit te modelit arrita ne saktësi te madhe deri ne 77%.

**Konkluzionet**: Prej 296 predikimeve ne total(unacc) 273 jan cilesuar sit të pa pranueshme, 21 si te pranueshme 2 si të mira dhe 0 si shum të mira. Prej 56 predikimeve për makinat e pranueshme(acc) 18 jan cilesuar sit e pa pranueshme , 32 jan cilesuar sit e pranueshme , 6 jan cilesuar te mira, dhe asnje si shum e mire. Prej 11 predikimeve të caktuara sit e mira(good) , asnje prej tyre nuk eshte cilesuar si unacc , 7 prej tyre jan cilesuar si acc , 4 prej tyre jan caktuar si good dhe asnje sis hum e mire vgood . Ne fund prej 69 predikimeve te cakuara si shum te mira(vgood) 5 prej tyre jan caktuar si unacc , 39 prej tyre jan cilesuar si acc , 9 prej tyre jan cilesuar sit e mira(good) dhe 16 prej tyre sis hum te mira vgood.

**Fjalët kyçe**

makinat · dyer · vleresimi · papranueshme · pranueshme

# HYRJE

Vleresimi I Makinave

Vlerësimi i kushteve të një makine përpara blerjes luan një rol vendimtar në vendimmarrje. Me ardhjen ne përfundim për vendimin e blerjes , klasifikimi i një makine me gjendje të mirë ose të pranueshme nga një makinë me kushte të papranueshme të kushtëzuara kërkon kohë dhe kërkon punë intensive. Ne mund të shfrytëzojmë teknikat e Mësimit të Makinerisë për të zhvilluar një sistem automatik për vlerësimin e makinave pasi ML ka treguar rezultate premtuese në zgjidhjen e problemeve të lidhura me klasifikimin. Në këtë projekt, ne do të analizojmë kualifikime të ndryshme fizike të një makine dhe më pas, do të ndihmojmë/rekomandojmë një përdorues në procesin e tyre të vendimmarrjes bazuar në atributet e veturave.

Nëse një konsumator do duheshte te blente nje makine ai do duhej te shkonte ne secilen kompani te makinave per te parë shitjet , statistikat e asaj makine , a eshte e qendrueshme kjo makinë . A ja vlen ajo makinë per tu blere nese ajo makinë është pak e shitur , ose ka siguri te ulet. Per kete arsyje me ane te keti dataseti ne do te mund te tregojmë se a ja vlen per tu blerë një makinë , me ato specifika ne baze te shum konsumatoreve dhe blerjeve te tyre .

Nese nje makine ka siguri te ulët dhe poashtu ka pak dyer dhe nuk është makinë e pranueshme , normale që konsumatori nuk do ta blej ate makinë .

Makinat qe pranohen jane ato makina qe jan te blereshme shum nga konsumatoret qe kane një qendrueshmeri mesatare , qe kane 4 dyer , qe ka fellne(LugBot) mesatare ose te vogla dhe që e kan sigurine te madhe , ne baze te ketyre statistikave vleresohet se ajo makine vlen per ta blerë.

# PUNË TË NGJASHME

Ka shum punë te ngjashme ne baze te ketij problemi , dhe ne do ti cekim shum prej tyre , me lloje te ndryshme te algoritmeve si :

Car Decision Tree Practice with Car Evaluation Dataset

By:MD. Abdullah Al Mamun

Car Evaluation | Random Forest ,Decision Tree , SVM

By:Narmin Humbatli

Car Evaluation with SVM and GridSearch

By:Pavlo Mospan

# METODOLOGJIA

Në këtë artikull, unë do të aplikojmë qasjet e të mësuarit të makinerisë (dhe përfundimisht do t'i krahasojmë ato) për të klasifikuar nëse një veturë është e pranueshme, e pa pranueshme, në gjendje të mirë ose në gjendje shum te mirë.

## Grupi i të dhënave

Grupi i të dhënave që kam përdorur është një csv-file (Comma Seperatted Values) i cili përmbanë 1729 të dhëna (vetura) si dhe klasifikimin për secilën veturë se a është e pranueshme,e prapranueshme, në gjendje të mire, në gjendje shumë të mirë.

## Paraprocesimi i të dhënave

Së pari i kam instaluar libraritë e nevojshme për të ekzekutuar kodin e python duke përdorur Algoritmin e Naïve Bayes.

Pastaj e kam importuar fajllin e Dataset-it dhe kemi filluar ta përdorim. Gjatë përdorimit kam hasur në vlera boshe ose te keq-shkruara dhe i kemi eleminuar menjëherë. Pas kësaj kam filluar që në kod të i ndajmë pjesët e fajllit qe duhet ti përdorim pra e kemi marrur vetëm kolonën e “Evaluation” që definonte se një veturë ishte e pa pranueshme, e pranueshme, në gjendje të mire ose ne gjendje shumë të mirë dhe i kemi marrur të dhënat e veturës nga kolonat e mbetura per të ardhur në përfundim.

## Modeli i trajnimit

Modeli i trajnimit është ndërtuar duke u bazuar në algoritmin e Naïve Bayes. Kodin e gatshëm e kam marrë në internet dhe e kam përshtatur për nevojat dhe kërkesat e detyrës .

Për fillim i kemi importuar këto paketa të python për të na ndihmuar në ekzekutimin e kodit dhe shfaqjen e rezultateve:

1. “Pandas” - për lexim të fajllit

2. “Nltk” - biblioteka të ndryshme të përpunimit të tekstit me shumë grupe të dhënash testuese.

3. “Seaborn” për paraqitjen vizuale të statistikave

# REZULTATET E FITUARA

Për të përshkruar rezultatët e fituara të klasifikimit të makinave nga dataset-i që kemi trajtuar, përdoret një matricë konfuzioni e cila mund të quhet si përmbledhje e të dhënave të parashikuara dhe aktuale.

Gjenerimi i matricës së konfuzionit specifikon shkallën e vërtetë pozitive, shkallën e vërtetë negative, shkallën e rreme positive dhe atë të rreme negative. …

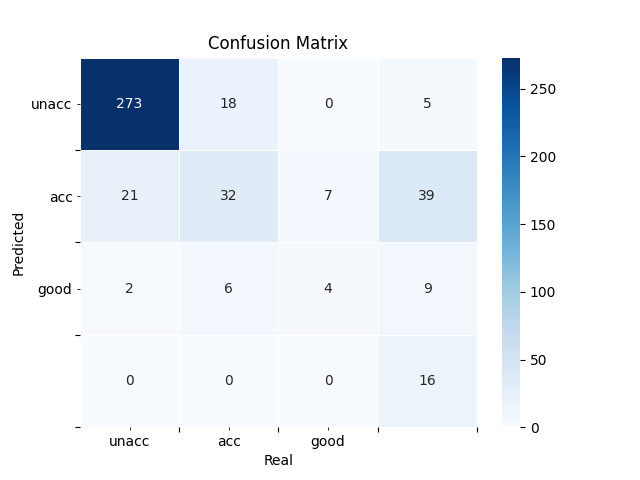


Figure 1.Confusion Matrix per dataset I cili permban 1729 rreshta ku bosti x paraqet vlerat reale te datasetit ndersa boshti y jan predikimet.

Confuxion Matrix te figura 1. , e paraqet matricen e fituar nga dataseti I klasifikuar.

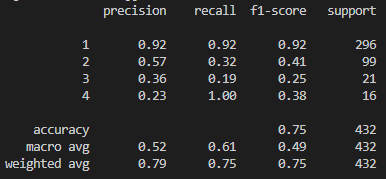
Sipas kolonës se parë prej 296 vleresimeve totale , 273 jan supozime te sakta qe janë të papranueshme 21 prej tyre jan cilësuar si të pranueshme , 2 prej tyre jan cilësuar sit e mira , dhe asnje ose 0 jan cilesuar si shumë të mira (vgood).

Ne kolonen e dyte prej 56 vleresimeve në total që jan klasifikuar sit e pranueshme , 18 prej tyre jan cilesuar si të pa pranueshme , 32 prej tyre jan cilësuar si pranueshme , 6 prej tyre jan cilësuar si të mira , dhe 0 jan cilesuar sis hum te mira .

Ne kololen e tretë prej 11 klasifikimeve të makinave të cilësuara si të mira , asnje nuk është e cilësuar si e papranueshme , 7 jan cilësuar si te pranueshme , 4 jan cilësuar si të mira dhe asnje nuk është cekur si shum e mir (vgood).

Ne Kolonen e 4 prej 69 klasifikimeve në total qe cilësohen si shum të mira . 5 prej tyre janë cilësuar si të pa pranueshme , 39 prej tyre jan cilesuar si të pranueshme , 9 prej tyre jan cilësuar si të mira dhe 16 prej tyre jan cilësuar si shum të mira.

Table 1. Conclusion Matrix . Klasifikimi I makinave per precision , recall, f1-score dhe support



Conslusion matrix – paraqet tabelen e rezultatit te datasetit me 1729 rreshta .

Ne Vleren 1 janë makinat(unaccepted) preciziteti eshte 0.92(92%)

Ne Vleren 2 janë makinat(accepted) ku preciziteti I saj eshte 0.57(57%)

Ne Vleren 3 janë makinat(good) ku preciziteti I saj eshte 0.36(36%)

Ne Vleren 4 janë makinat(vgood) ku preciziteti I saj eshte 0.23(23%)

Ndërsa sipas Confusion Matrix kemi llogaritur precision, recall dhe f1-score .

* Precision (Preciziteti) – përcakton se sa në të vërtetë te dhënat janë pozitive nga të gjitha pozitivet e parashikuara.
* Recall - Nga totali pozitiv, përcakton sa përqind janë parashikuar pozitive. Është njësoj si TPR (shkalla e vërtetë pozitive).
* F1-Score – Merr parasysh si pozitivet e rreme ashtu edhe negativet e rreme. Prandaj, funksionon mirë në një grup të dhënash të çekuilibruar.
* Support – Eshte totali I predikimeve per attribute.

# KONKLUZION

Njerzit nëpër botë blejnë makina të ndryshme , dhe blerjet e tyre jan të shpeshta .Për këtë arsyje edhe klasifikimi i këti dataseti është me rëndësi . Sepse njerzit shpesh hërë edhe nuk din se qfar makine të blejnë dhe në nxitim e sipër mund të blejnë makinën e gabuar dhe pastaj ajo makinë me ato cilësi nuk i përshtatet . Në këtë trajnim të datasetit kam arritur ti klasifikoj makinat që janë të pranueshme për tu blerë , të pa pranueshme , të cilësuara si të mira dhe shum të mira .Duke e trajnuar modelin me alogitmin Naïve Bayess .Ku saktesia e tij ka shkuar deri ne 77%

REFERENCAT

1. Mohammad Masum ,(2022).Car Evaluation Analysis Using Decision Tree Clasifier

[http://towardsdatascience.com/car-evaluation-analysis-using-desicion-tree-clasifier-61a8ff12bf6f](http://towardsdatascience.com/car-evaluation-analysis-using-desicion-tree-clasifier-61a8ff12bf6f%20)

1. Ahiale Darlington ,(2018). Car Evaluation Dataset

<https://www.kaggle.com/datasets/elikplim/car-evaluation-data-set>

1. John Mantios, (2020). Car Evaluation dataset clasification

<https://www.kaggle.com/johnmantios/car-evaluation-dataset-clasification/>

1. John Mantios, (2020). Car Evaluation Data Set

<https://www.kaggle.com/johnmantios/car-evaluation-dataset-clasification/>

1. **UKEssays. (November 2018). Car Evaluation Using Machine Learning. Retrieved from**

<https://www.ukessays.com/essays/computer-science/car-evaluation-using-machine-learning.php?vref=1>