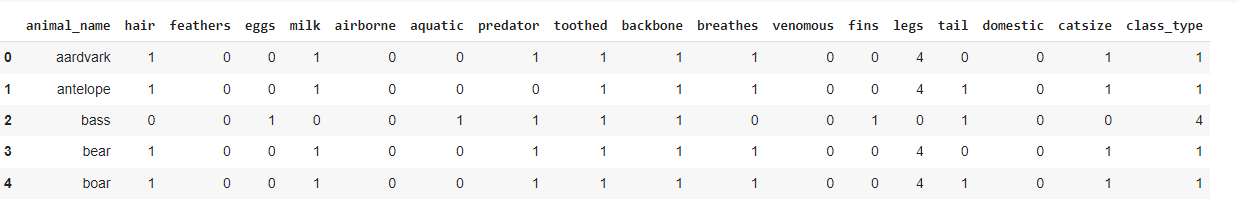
**Machine iLearning iModel i iFor iClassification iof iZoo iAnimals ibased ion iAnimal ikingdoms**

Affiliation i: **iLAGISHETTY iNANDITHA**

Correspondence: i2203a52097@sru.edu.in; iSection: iAIML-AA i

SR iUniversity-school iof ics i&ai.

|  |  |
| --- | --- |
| **Abstract: i**The i izoo ianimal iclassification idataset itaken ifrom iKaggle i, ia icollection iof ifeatures iincludes:  animal\_name: iUnique ifor ieach iinstance, ihair iBoolean i,feathers iBoolean, ieggs iBoolean, i  milk iBoolean, iairborne iBoolean i,aquatic iBoolean, ipredator iBoolean, itoothed iBoolean i,backbone iBoolean, ibreathes iBoolean, ivenomous iBoolean, ifins iBoolean, ilegs iNumeric i(set iof ivalues: i{0,2,4,5,6,8}),tail iBoolean i,domestic iBoolean i,catsize iBoolean i,class\_type iNumeric i(integer ivalues iin irange i[1,7]) |  |
| , ioffers iunique ichallenges ifor iaccurate iclassification. iWith ithis icontext, | 2 |
| our iresearch iaims ito iadvance ithe ifield’s iunderstanding iof ithe ibest imodel iselection istrategy iwhile | 3 |
| taking iinto iaccount ithe ispecific icharacteristics iof ithis idataset. | 4 |
| In ithe iexperimental iphase, iwe imeticulously ipreprocess ithe idata, iaddressing iissues isuch ias imissing | 5 |
| values iand ifeature iscaling. iThe iselected imachine ilearning imodels iundergo irigorous itraining iand | 6 |
| evaluation, iconsidering istandard iperformance imetrics iincluding iaccuracy, iprecision, irecall, iand | 7 |
| F1-score. iTo iprovide iinsights iinto imodel istability, iwe iemploy ibootstrap iresampling, ia itechnique ithat | 8 |
| enables ius ito iassess ivariations iin imodel iperformance. | 9 |
| **Keywords: i**Support iVector iMachines i(SVM), iK-Nearest iNeighbors, iconfusion imatrix i(heat imaps), | 10 |
| (KNN), iData iPreprocessing, iBootstrapping, iModel iEvaluation, iAccuracy, iPrecision, iRecall, iF1Score, iModel | 11 |
| Selection, iModel iStability, iFeature iEngineering, iComparative iAnalysis, iResearch iMethodology, iModel | 12 |
| **1. iIntroduction** | 13 |
| Zoo ianimal iclassification idataset iwas itaken ifrom iUGI iMACHINE iLEARNING i. iThe idata iset icontains idetails iof imany ianimals. iBased ion ithese idetails iwe iconclude iwhich ianimal ibelongs ito iwhich ianimal ikingdom. iThere iare i16 ivariables iwith ivarious itraits ito idescribe ithe ianimals.  The i7 iClass iTypes iare: iMammal, iBird, iReptile, iFish, iAmphibian, iBug iand iInvertebrate.  This imodel ican ibe iused iin iscientific ifields iwhere ianimals ishould ibe iclassified. iScientists iwho idiscovers inew ianimals ican iuse ithis imodel ito ifind iout ithe iclass iand ikingdom iof ithat ianimal ibased ion ithese ifeatures.  This ihelps ius iunderstand ihow ithe ispecies iare irelated ito ieach iother. ithis iallows ius ito ibetter igrasp itheir iqualities iand idistinctions ifrom iother iorganisms.    **Figure i1. i**Animal iclassification. |  |
| **2. iLiterature iReview** | |
| The ifollowing iliterature ireview iprovides ian ioverview iof ikey iresearch iareas iin ithe | |
| context iof ithe iZOO iANIMAL iCLASSIFICATION iand ithe iapplication iof imachine ilearning ifor ithe | |
| classification iof ianimals. | |
| *2.1. iMachine iLearning iin izoo ianimal iclassification* | |
| Machine ilearning ihas iincreasingly ibecome ia ivaluable itool iin i*zoo ianimal iclassification idataset* | |
| for ithe iclassification iand ianalysis iof ianimals. iThe iunique icapabilities iof ithe idataset | |
| ihave iled ito ia igrowing ibody iof iresearch ithat iemploys imachine ilearning itechniques | |
| to iclassify ianimals. | |
| *2.2. iPrevious iStudies iin izoo ianimal iclassification* | |
| Several inotable istudies ihave iexplored ithe iclassification iof ianimals iusing | |
| data icollected iby ithe iUGI iMACHINE iLEARNING. iFor iexample, iMIRI iCHOI iused iArtificial ineural iNetwork(ANN),SLP(Single iLayer iPerceptron) | |
| ito idistinguish ibetween ivarious iclasses iof ianimals | |
|  | |
| *2.3. iChallenges iand iResearch iGaps* | |
| While ithese istudies iprovide ivaluable iinsights iinto ithe iapplication iof imachine ilearning | |
| in iclassification iof ianimals, icertain iresearch igaps ipersist. iMost | |
| notably, ithere iis ia ilimited iexploration iof ithe icombined iapplication iof imultiple imachine | |
| learning imodels, isuch ias iknn i, iand iSVM, iin ithe icontext iof | |
| Zoo ianimal iclassification. iFurthermore, ithere iis ia ineed ifor ian iextensive iexami- | |
| nation iof ifeature iengineering itechniques iand idata ipreprocessing istrategies iin ioptimizing | |
| classification iaccuracy. | |
| **3. iData iand iMethodology** | |
| *3.1. iData iDescription* | |
| As idocumented iby iRODRIGO iHJORT, ithere iare i101 irows iin ithis | |
| data iset; i48 iof iwhich iare iMammals,32 iof iwhich iare ibirds,26 iof iwhich iare ifish,21 i iof iwhich iare iinvertebre,18 iof iwhich iare ibugs,15 iof iwhich iare iamphibians iand i15 iof iwhich iare ireptiles. | |
| The itarget iis iimbalanced imost iof irecords ibeing imammals. ithe idata iprovided iat ithe iUCI ihosting isite | |



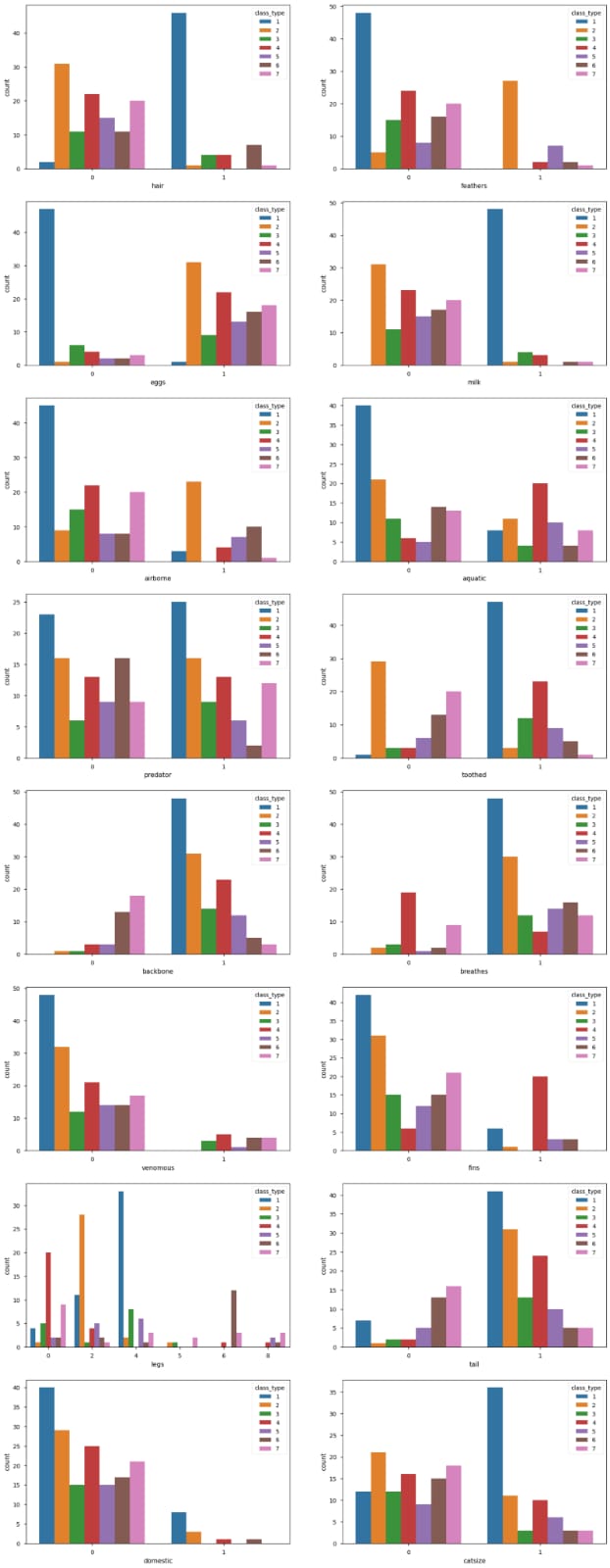
**Figure i2. i**Sample itraining idata

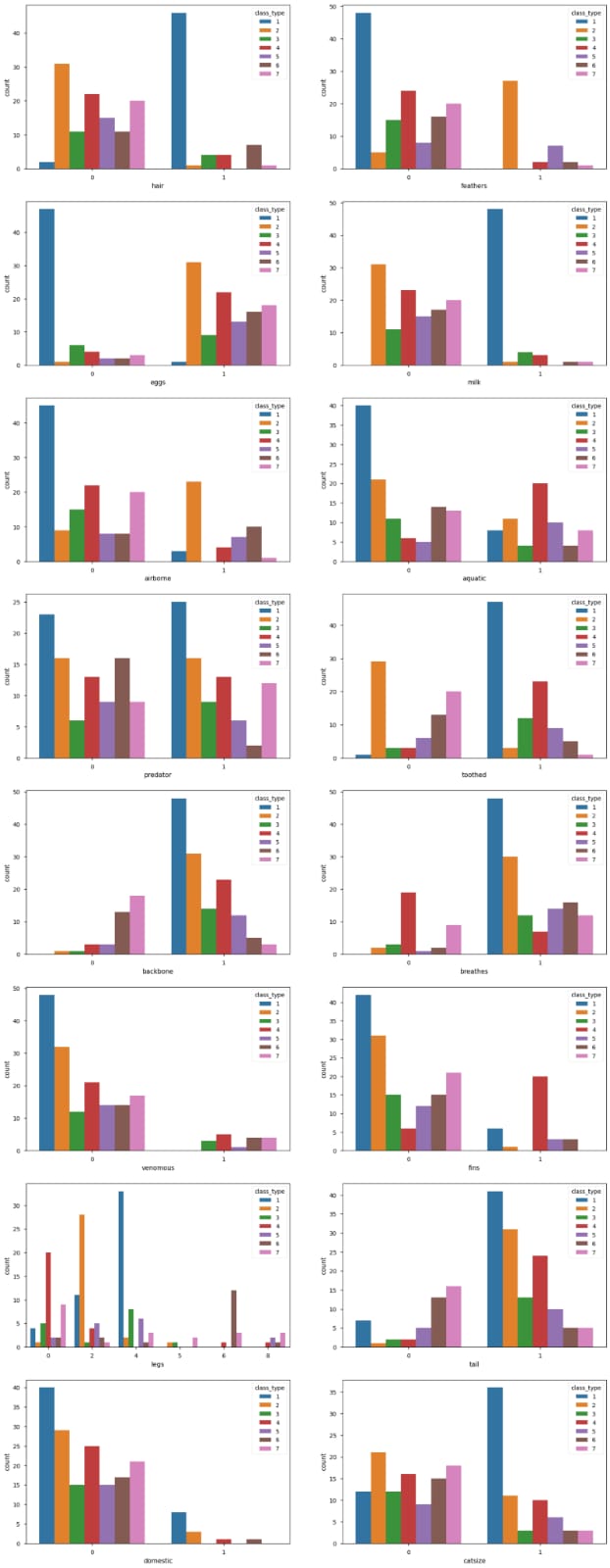
as ithe inumber iof irows iare ivery iless iI iapplied ismote imethod ito iincrease ithe inumber iof irows

*3.2. iData iAnalysis*

These igraphs ihelp iyou iunderstand ihow iindividual ifeatures icontribute ito ithe iclassification iproblem. iFeatures iwith idistinct idistributions ifor ithe itwo iclasses iare itypically imore iinformative ifor iclassification.

By ianalyzing ithese i, iyou ican imake iinformed idecisions iabout ifeature iselection, imodel ichoice, iand ifeature iengineering ito iimprove ithe iperformance iof iyour iclassification imodel.

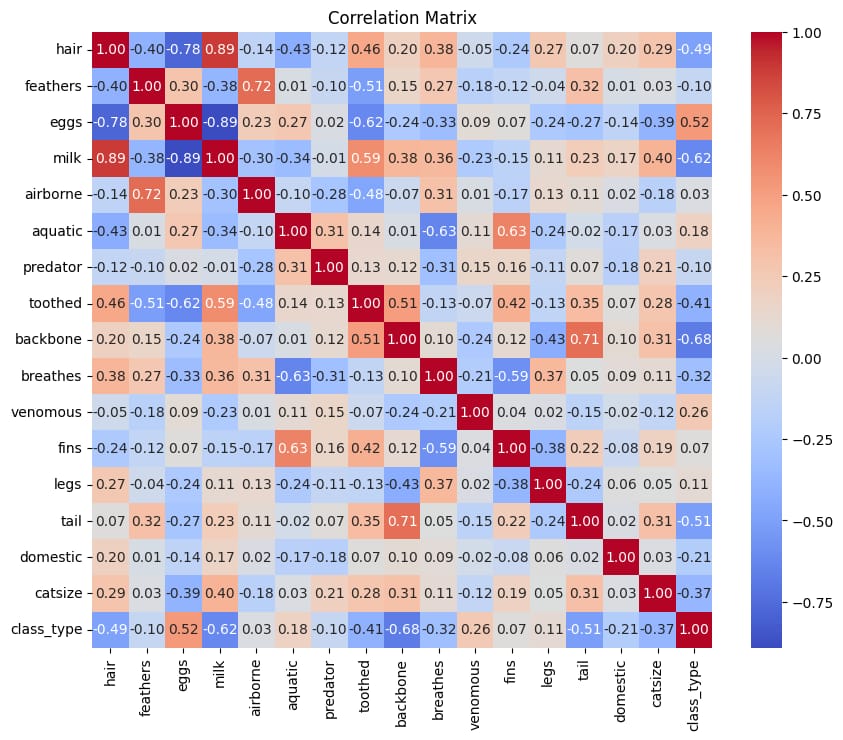




**Figure i3. i**probability ivs ifeature igraph

3.2.1. iCorrelation iMatrix

A icorrelation imatrix iis ia itable ithat ishows ithe irelationships ibetween imultiple ivariables iin ia idataset.



**Figure i4. i**Correlation imatrix

strength iand idirection iof ithe ilinear irelationship ibetween ipairs iof ivariables. iHere’s ia ibrief iexplanation iof ia icorrelation imatrix

1. Positive iCorrelation i(0 ito i1): iVariables imove iin ithe isame idirection iWhen ione iincreases, ithe iother itends ito iincrease ias iwell
2. Negative iCorrelation i(-1 ito i0): iVariables imove iin iopposite idirections. iWhen ione iincreases, ithe iother itends ito idecrease.
3. No iCorrelation i(0): iThere iis ino ilinear irelationship ibetween ithe ivariables.
   1. *Data iPreprocessing*

Data ipreprocessing iis ia icritical istep iin imachine ilearning ithat iinvolves icleaning, itrans iforming, iand iorganizing iraw idata iinto ia iformat isuitable ifor imodel itraining. iIt iplays ia isignificant irole iin iensuring ithat ithe idata iis iof ihigh iquality iand ithat ithe imachine ilearning imodel ican ilearn imeaningful ipatterns. iHere’s ian ielaborate iexplanation iof ivarious iaspects iof idata ipreprocessing:

|  |
| --- |
| 1. **Data iCleaning:** |
| *Handling iMissing iValues: i*Identify iand ihandle imissing idata, ieither iby iremoving irows |
| or ifilling iin imissing ivalues iusing itechniques ilike imean, imedian, ior iinterpolation. iThere |
| are ino inull ivalues iin ithis idataset |
| 2. **Data iTransformation:** |
| *Feature iScaling: i*Normalize ior istandardize inumerical ifeatures ito iensure ithat idifferent |
| features iare ion ia isimilar iscale. iCommon itechniques iinclude iMin-Max iscaling iand |
| **3.Data iSplitting:** |
| *Train-Validation-Test iSplit: i*Divide ithe idataset iinto itraining, ivalidation, iand itest isets. |
| The itraining iset iis iused ito itrain ithe imodel, ithe ivalidation iset iis iused ito itune ihyperpa- |
| rameters, iand ithe itest iset iis iused ito ievaluate imodel iperformance. |
| **4.Handling iString iData:** |
| *Text iPreprocessing: i*For inatural ilanguage iprocessing itasks, ipreprocess itext idata iby |
| tokenizing, iremoving istop iwords, istemming ior ilemmatizing, iand iconverting itext ito |
| numerical irepresentations i(e.g., iTF-IDF ior iword iembeddings). |

Model iSelection iWe iselected itwo imachine ilearning imodels ifor iour ianalysis:

Support iVector iMachines iand iK-Nearest iNeighbors. iThese imodels 118 iwere ichosen ibased ion itheir isuitability ifor imultiple iclass i iclassification itasks. iPerceptron iand ilogistic imethods iare inot iused ihere ias ithey iare iunfit ifor imultiple iclass iclassification

*3.4. iExperimental iSetup*

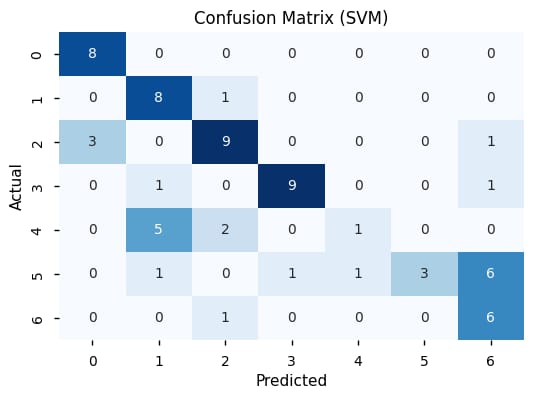
Train i i i i itest isplitting i icode iperforms ithe iactual idata isplit. iIt itakes ithe ishuffled iData iFrame, iwhich icontains iall ithe idata, iand isplits iit iinto itwo iseparate iData iFrames: ithe itraining iset iand ithe itest iset. iThe i iproportions ican ibe iadjusted ibased ion ithe ispecific irequirements iof iyour imachine ilearning i itask

**4.models**

*4 i.1 iSUPPORT iVECTOR iMACHINE*

The imodel’s iperformance iwas iassessed iusing iprecision, irecall, iand iF1-score imetrics, i132 ialong iwith ian iaccuracy iscore. iThe idataset iconsisted iof i174 isamples, idivided iinto i7classes: iClass i1 i- iClass i7. iThe imodel iachieved ian ioverall iaccuracy iof i64 ipercent, iindicating ihas ipredictive icapability. iClass-specific iperformance imetrics irevealed iimpressive i iprecision, irecall, iand iF1-scores, ifurther iunderscoring ithe imodel’s ieffectiveness. i136 iAccuracy: iThe imodel idemonstrated ian ioverall iaccuracy iof i64 ipercent, isignifying iits iability i137 ito icorrectly iclassify ithe imajority iof iinstances.

1. Precision: iThe iprecision ifor iClass i0 iis i99 ipercent, iindicating ithat ialmost iall ipredicted iClass i0 iinstances iwere iindeed iClass i0.
2. Recall: iThe irecall ifor iClass i0 iis i99 ipercent, isuggesting ithat inearly iall iactual iClass i0 iinstances iwere icorrectly iidentified.
3. F1-Score: iThe iF1-score ifor iClass i0 iis i99 ipercent, irepresenting ia ibalanced iharmonic imean iof iprecision iand irecall ifor iClass i0.



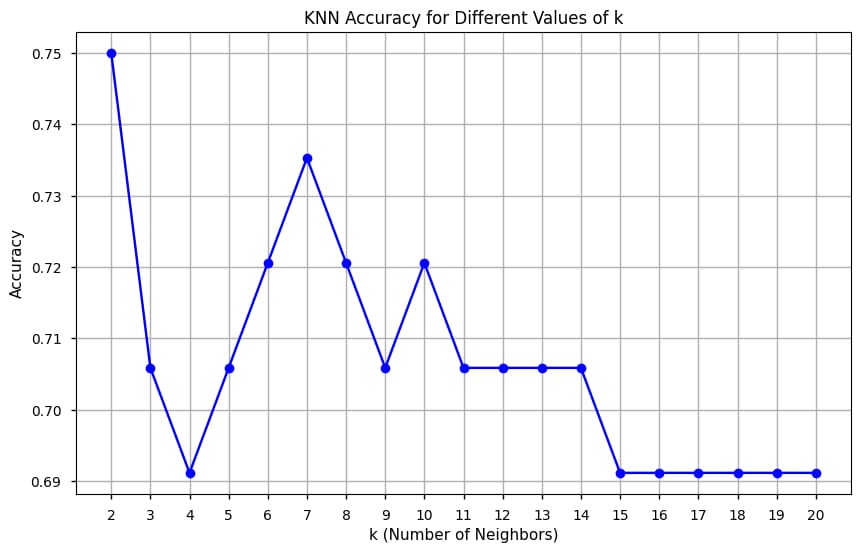
|  |
| --- |
| **Figure i5. i**Confusion iMatrix |
| It’s ievident ifrom ithe iconfusion imatrix ithat: |
| **True iNegatives i(TN): i**There iare i18 iinstances ifor iwhich iboth ithe iactual iclass iand ithe |
| Support ivector imachine imodel’s iprediction iare inegative i(0). iThese iare icorrectly iclassified iinstances |
| where ithe imodel iaccurately iidentified ithe inegative iclass. |
| **True iPositives i(TP): i**There iare i44 iinstances ifor iwhich iboth ithe iactual iclass iand ithe |
| model’s iprediction iare ipositive i(1). iThese iare icorrectly iclassified iinstances iwhere ithe imodel |
| accurately iidentified ithe ipositive iclass. iThe imodel’s iperformance iwas iassessed iusing iprecision, irecall, iand iF1-score imetrics, i243 ialong iwith ian iaccuracy iscore. iThe idataset iconsisted iof i174 isamples, idivided iinto itwo iclasses: iClass i1 i-Class i7. iThe imodel iachieved ian ioverall iaccuracy iof i67 ipercent, iindicating iits ihigh ipredictive icapability. iClass-specific iperformance imetrics irevealed iimpressive i i i i iprecision, irecall, iand iF1-scores, ifurther iunderscoring ithe imodel’s ieffectiveness. |

Accuracy: iThe imodel idemonstrated ian ioverall iaccuracy iof i i67 ipercent, isignifying iits iability i248 ito icorrectly iclassify ithe imajority iof iinstances

1. iPrecision: iThe iprecision ifor iClass i0 iis i95 ipercent, iindicating ithat ialmost iall ipredicted iClass i0 iinstances iwere iindeed iClass i0.

2. iRecall: iThe irecall ifor iClass i0 iis i97 ipercent, isuggesting ithat inearly iall iactual iClass i0 i254 iinstances iwere icorrectly iidentified.

3. iF1-Score: iThe iF1-score ifor iClass i0 iis i96 ipercent, irepresenting ia ibalanced iharmonic imean iof iprecision iand irecall ifor iClass i0



**Figure i6.**K-NEAREST iNEIGHBOUR iMAP

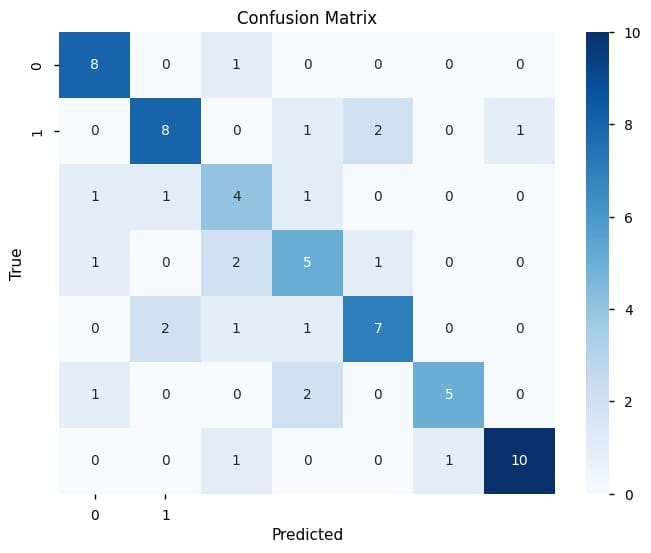


Fig7: iconfusion imatrix.

|  |
| --- |
| It’s ievident ifrom ithe iconfusion imatrix ithat: |
| **True iNegatives i(TN): i**There iare i21 iinstances ifor iwhich iboth ithe iactual iclass iand ithe |
| Perceptron imodel’s iprediction iare inegative i(0). iThese iare icorrectly iclassified iinstances |
| where ithe imodel iaccurately iidentified ithe inegative iclass. |
| **True iPositives i(TP): i**There iare i47 iinstances ifor iwhich iboth ithe iactual iclass iand ithe |
| model’s iprediction iare ipositive i(1). iThese iare icorrectly iclassified iinstances iwhere ithe imodel |
| accurately iidentified ithe ipositive iclass.  *4.5. iBootstrap imethod* |
| Bootstrapping iis ia itechnique iused ifor iresampling iuually iused iin imachine ilearning iand |
| Statistical ilearning. iBootstrapping igenerally isample idata irepeat ifrom iyour idataset iwith ireplacement i |
| To icreate imany inew idatasets, ievery idata iwill icontain ithe iexact iequal isize ias ithat iof ithe ioriginal idata iset i. ithe imain iaim iof ibootstrapping iis ito iform imany i i inew idatasets i |
| These idatasets iare icalled i"bootstrap isamples. iBoot istrapping iis iuseful ito i iassess |
| the idisparity iand ivalidity iof iyour imodel. iWe ican iuse imany imodels ialong iwith iboot istrapping iand iget ito ia iconclusion ihow iour ibootstrap isamples iworks ion idifferent isamples iof idataset. iThe ibootstrapping iis iused ito iestimate iconfidence iintervals ifor ivarious imeasures iof imodel iperformance, isuch ias ithe itrue ior isquared ierror. iResampling idata iiteratively |
| **Confidence iIntervals ifor isupport ivector imachine:** |
| *95.0 iConfidence iInterval: i*This iinterval iis iused ito iestimate ia irange iwithin iwhich iyou ican |
| be i95 ipercent iconfident ithat ithe itrue iaccuracy iof iyour isupport ivector imachine imodel ifalls. |
| *0.15 iConfidence iInterval i(Lower iBound): i*This irepresents ithe ilower iend iof ia irange iwithin |
| which iyou ican ibe i98.6 ipercent iconfident ithat ithe itrue iaccuracy iof imodel ifalls. |
| *0.31 iConfidence iInterval i(Upper iBound): i*This irepresents ithe iupper iend iof ia irange iwithin |
| which iyou ican ibe i99.8 ipercent iconfident ithat ithe itrue iaccuracy iof imodel ifalls.   |  |  | | --- | --- | | *MEAN iACCURACY(BOOTSTRAPPED)* | *0.15* | | *95% iCONFIDENCE iINTERVAL* | *0.01-0.31* | | *COEFFICIENT iOF iVARIATION i* | *57.14%* | | *STANDARD iDEVIATION iOF iACCURACY* | *0.08* | | *MEAN iSQUARE iERROR* | *0.01* | |

**Table i1:BOOTSTRAPPING iFOR iSVM i(SUPPORT iVECTOR iMACHINE)**

**Confidence iIntervals ifor ik-nearest ineighbours:**

|  |
| --- |
| *95.0 iConfidence iInterval: i*This iinterval iis iused ito iestimate ia irange iwithin iwhich iyou ican |
| be i95 ipercent iconfident ithat ithe itrue iaccuracy iof iyour ik inearest ineighbour imodel ifalls. |
| *0.15 iConfidence iInterval i(Lower iBound): i*This irepresents ithe ilower iend iof ia irange iwithin |
| which iyou ican ibe i98.6 ipercent iconfident ithat ithe itrue iaccuracy iof imodel ifalls. |
| |  |  | | --- | --- | | *MEAN iACCURACY(BOOTSTRAPPED)* | *0.04* | | *95% iCONFIDENCE iINTERVAL* | *0.04-0.04* | | *COEFFICIENT iOF iVARIATION i* | *0.00%* | | *STANDARD iDEVIATION iOF iACCURACY* | *0.00* | | *MEAN iSQUARE iERROR* | *0.00* | |
| **TABLE i2: iBOOTSTRAPPING iFOR iKNN i(K- iNEAREST iNEIGHBOURS)**  *0.31 iConfidence iInterval i(Upper iBound): i*This irepresents ithe iupper iend iof ia irange iwithin |
| **KNN iClassifier** |
| 95.0 iconfidence iinterval i0.04(lower) iand i0.04(higher) |
| Mean iAccuracy(KNN): i0.04 |
| **Mean iAccuracy i(KNN): i**0.04 |
| The imean iaccuracy iof iKNN iis iexceptionally ihigh, iat i95 ipercent. iThis iindicates ithat ithe |
| model iis inot iso iaccurate iand ipredicts ioutcomes iwith ia ilow idegree iof isuccess. |
| **Standard iDeviation i(KNN): i**0.00 |
| The istandard ideviation imeasures ithe ivariability ior ispread iof idata ipoints. iA istandard |
| deviation iof i0.00 imeans ithat ithe iaccuracy iscores iof iKNN iare iextremely iconsistent iand |
| do inot ivary imuch. iThis ihigh iconsistency isuggests ithat ithe imodel’s iperformance iis |
| **Confidence iIntervals:** |
| *95.0 iConfidence iInterval: i*This iinterval iis iused ito iestimate ia irange iwithin iwhich iyou ican |
| be i95 ipercent iconfident ithat ithe itrue iaccuracy iof iKNN ifalls. |
| *0.04 iConfidence iInterval i(Lower iBound): i*This irepresents ithe ilower iend iof ia irange iwithin |
| which iyou ican ibe i98.6 ipercent iconfident ithat ithe itrue iaccuracy iof imodel ifalls. |
| *0.04 iConfidence iInterval i(Upper iBound): i*This irepresents ithe iupper iend iof ia irange iwithin |
| which iyou ican ibe i99.8 ipercent iconfident ithat ithe itrue iaccuracy iof imodel ifalls. |

**Pca i(principal icomponent ianalysis)**

Principal icomponent ianalysis iis ia imethod iused ifor ireducing idimentionality.

If iour idata iset icontains imore ithan i10 icolumns ithen iwe ican iapply ithis imethod ito ireduce ithe inumber iof irows.

1. Before iprincipal icomponent ianalysis ithe iaccuracy iis i0.64 iin isupport ivector imachine imodel iand iafter iapplying iprincipal icomponent ianalysis ithe iaccuracy iis i0.5. i.
2. Before iprincipal icomponent ianalysis ithe iaccuracy iis i0.75 iin ik inearest ineighbour imodel i iand iafter iapplying iprincipal icomponent ianalysis ithe iaccuracy iis i0.7
3. **Conclusion:**

The ifinal iconclusion iof ithis iproject iis ithat iwe iare iable ito ifind ia igood iaccuracy imodel iand iwe iare iable ito iclassify ianimals ibased ion ithe ifeatures iit icontain. iMy ifuture iplans ifor ithis iproject iwill ibe:

Including ia iinterface ifor ieasy iuse i i. iproviding ian iapplication iso ithat iit ican ibe iused iby iscientists i. iI i iam ithinking iof i imodel ithat ican iclassify ianimal ibased ion ithe iinput iphoto igiven i. ireading ithe iphoto iit ishould irecognise iall ithe ifeatures iand ithen igive ioutput.

****

**Figure i8: iconclusion iof ianimal iclassification.**

|  |  |  |
| --- | --- | --- |
| **6. iReferences** | | |
| 1. | Zoo iAnimal iClassification | |
|  | Use iMachine iLearning iMethods ito iCorrectly iClassify iAnimals iBased iUpon iAttributes  <https://www.kaggle.com/datasets/uciml/zoo-animal-classification>-**UCI iMACHINE iLEARNING** | |
| 2. | ANN/SLP] iMaking iModel ifor iMulti-Classification i-<https://www.kaggle.com/code/mirichoi0218/ann-slp-making-model-for-multi-classification>MIRI iCHOI i-**kaggle** | |
| 3. | <https://pranavkumar623.medium.com/zoo-animal-classification-e4cc070a5d44-pranav> ikumar.-**medium** | |
| 4. | <https://github.com/sharmaroshan/Zoo-Dataset>sharmaroshan-**github** | |
| **Disclaimer/Publisher’s iNote: i**The istatements, iopinions iand idata icontained iin iall ipublications iare | |
| solely ithose iof ithe iindividual iauthor(s) iand icontributor(s) iand inot iof iMDPI iand/or ithe ieditor(s). | |
|  | |
| MDPI iand/or ithe ieditor(s) idisclaim iresponsibility ifor iany iinjury ito ipeople ior iproperty iresulting ifrom | |
| any iideas, imethods, iinstructions ior iproducts ireferred ito iin ithe icontent. | |