**Project Title: “Species Distribution Modeling”**

**Subject: Data Mining and Data Warehousing**

**Group Members: Khwaja Zain – ur – Rehman (EP 20101020)**

**Ali Ilyas (EP 20101007)**

**Hamza Hafeez (EP 20101016)**

**Muhammad Areeb (EP 20101024)**

**Degree Name: MCS (Masters in Computer Science)**

**Year: 2021**

**Department Name: UBIT (Umaer Basha Institute of Technology)**

**University Name: UOK (University of Karachi)**

**Table of Contents**

**1.** **Abstract**  **(03)**

**2.** **Introduction** **(04)**

**2.1** Niche-based models from presence-only data **----------------------------------------------------------- (06)**

**2.2** Existing approaches for presence-only modeling **------------------------------------------------------- (07)**

**2.3** Maxent **-------------------------------------------------------------------------------------------------------- (07)**

**3. Methods (08)**

**3.1** Maxent details **----------------------------------------------------------------------------------------------- (08)**

**3.2** A machine learning perspective **--------------------------------------------------------------------------- (08)**

**3.3** A Maxent implementation for modeling species distributions **---------------------------------------- (09)**

**3.4** GARP **--------------------------------------------------------------------------------------------------------- (09)**

**3.5** Data sources **-------------------------------------------------------------------------------------------------- (10)**

**3.6** Environmental variables **------------------------------------------------------------------------------------ (10)**

**3.7** Model building **---------------------------------------------------------------------------------------------- (11)**

**3.8** Model evaluation **-------------------------------------------------------------------------------------------- (12)**

**4. Results (12)**

**4.1** Quantitative results **------------------------------------------------------------------------------------------ (12)**

**4.2** Visual interpretation **---------------------------------------------------------------------------------------- (15)**

**4.3** Models derived from climatic and elevational variables **----------------------------------------------- (15)**

**4.4** Addition of potential vegetation variable **---------------------------------------------------------------- (16)**

**5. Discussion and conclusions (17)**

**5.1** Statistical tests **----------------------------------------------------------------------------------------------- (17)**

**5.2** Biological interpretations **---------------------------------------------------------------------------------- (18)**

**5.3** Spatial context of errors **------------------------------------------------------------------------------------ (20)**

**5.4** Advantages of Maxent **-------------------------------------------------------------------------------------- (21)**

**5.5** Future work **-------------------------------------------------------------------------------------------------- (22)**

**“Species iDistribution iModeling”**

**1. iAbstract:-**

The ivacuity iof idetailed ienvironmental idata, itogether iwith iaffordable iand iimportant icomputers, ihas ifueled ia irapid-fire iincrease iin iprophetic imodeling iof ispecies ienvironmental iconditions iand igeographic idistributions. iFor isome ispecies, idetailed ipresence/ iabsence icircumstance idata iare iavailable, iallowing ithe iuse iof ia ivariety iof istandard istatistical iways. iStill, iabsence idata iaren't iavailable ifor iutmost ispecies. iIn ithis ipaper, iwe iintroduce ithe iuse iof ithe imaximum ientropy isystem i(Maxent) ifor imodeling ispecies igeographic idistributions iwith ipresence-only idata. iMaxent iis ia igeneral- ipurpose imachine iliteracy isystem iwith ia isimple iand iprecise ifine iexpression, iand iit ihas ia inumber iof iaspects ithat imake iit iwell- isuited ifor ispecies idistribution imodeling. iIn iorder ito iprobe ithe iefficacity iof ithe isystem, ithen iwe iperform ia iinternational-scale icase istudy iusing itwo iNeotropical imammals ia itableland ispecies iof iidleness, iBradypus ivariegatus, iand ia ismall imontane imurid irodent, iMicroryzomys iminutus. iWe icompared iMaxent iprognostications iwith ithose iof ia igenerally iused ipresence-only imodeling isystem, ithe iInheritable iAlgorithm ifor iRule- iSet iVaticination i(GARP). iWe imade iprognostications ion i10 iarbitrary isubsets iof ithe icircumstance irecords ifor iboth ispecies, iand ialso iused ithe iremaining ipoints ifor itesting. iBoth ialgorithms ihanded ireasonable iestimates iof ithe ispecies’ irange, ifar isuperior ito ithe ishadowed ifigure imaps iavailable iin ifield iattendants. iAll imodels iwere isignificantly ibetter ithan iarbitrary iin iboth ibinomial itests iof ielision iand ireceiver ioperating icharacteristic i(ROC) ianalyses. iThe iarea iunder ithe iROC iwind i(AUC) iwas inearly ialways iadvanced ifor iMaxent, iindicating ibetter idemarcation iof isuitable iversus iinfelicitous iareas ifor ithe ispecies. iThe iMaxent imodeling iapproach ican ibe iused iin iits ipresent iform ifor inumerous ioperations iwith ipresence-only idatasets, iand igraces ifarther iexploration iand idevelopment.**2.** i**Introduction:-**

Modeling ispecies’ igeographic idistributions iis ian iimportant iproblem iin iconservation ibiology. iIn ithis iexample iwe imodel ithe igeographic idistribution iof itwo isouth iamerican imammals igiven ipast iobservations iand i14 ienvironmental ivariables. iSince iwe ihave ionly ipositive iexamples i(there iare ino iunsuccessful iobservations), iwe icast ithis iproblem ias ia idensity iestimation iproblem iand iuse ithe i*OneClassSVM* ias iour imodeling itool. iThe idataset iis iprovided iby iPhillips iet. ial. i(2005). i*Species iDistribution iModeling. i* iNew iYork: iPhillips. iIf iavailable, ithe iexample iuses i*basemap* ito iplot ithe icoast ilines iand inational iboundaries iof iSouth iAmerica.

The itwo ispecies iare:

* “*Bradypus ivariegatus*” i, ithe iBrown-throated iSloth.
* “*Microryzomys iminutus*” i, ialso iknown ias ithe iForest iSmall iRice iRat, ia irodent ithat ilives iin iPeru, iColombia, iEcuador, iPeru, iand iVenezuela.

1. Bradypus ivariegatus:-
   * Justification:-

*Bradypus ivariegatus* iis ilisted ias iLeast iConcern iin iview iof iits iwide idistribution iincluding ia ilarge ipart iof ithe iAmazon iforest, ipresumed ilarge ipopulation, iits ioccurrence iin ia inumber iof iprotected iareas, iand ibecause iit iis iunlikely ito ibe ideclining ifast ienough ito iqualify ifor ilisting iin ia ithreatened icategory.

* + Geographic iRange iInformation:-

Bradypus ivariegatus iranges ifrom iHonduras iin ithe inorth, ithrough isouthern iCentral iAmerica. iIn iSouth iAmerica, iit iranges ifrom iColombia iinto iwestern iand isouthern iVenezuela, iand isouth iinto iEcuador, ieastern iPeru iand iBolivia, iinto iBrazil iand inorthern iArgentina i(where iit's inow iconsidered ito ibe iannihilated). iIts idistribution ioverlaps iwithB. itorquatus iin ithe icentral ipart iof ithe iAtlantic itimber i(Hirsch iand iChiarello i2012). iIn iBrazil, ithe ispecies ipresently ioccurs iin iforested iareas iof ithe iAmazon, iAtlantic itimber, iand iconceivably iin ithe icontact izones ibetween ithese ibiomes iand iCerrado. iThere iare iliteral irecords iofB. ivariegatus iin ithe iCaatinga ibiome i(Moraes-

Barros iunpublished idata i2010). iThere iare ino iverified irecords iforB. ivariegatus iin ithe iPantanal ibiome iof iBrazil, ibut ithe ispecies imight ido iin ithe icontact izones ibetween ithis ibiome iand ithe iAmazon itimber ito ithe inorth. iFresh ifield istudies iare inecessary iin iorder ito iduly idefine ithe icurrent ispecies idistribution iin ithe iCerrado, iCaatinga iand iPantanal.

* + Threats iInformation:-

It iappears ithat ithere iare ino imajor ithreats ito i*B. ivariegatus* iat ithe iglobal ilevel. iNevertheless, isome isubpopulations, iespecially iin iColombia iand ithe iAtlantic iForest iin iBrazil, iare ideclining idue ito ideforestation ileading ito isevere ihabitat idegradation iand ifragmentation. iThe ilowest ilevels iof igenetic idiversity iof ithe ispecies iwere iobserved iin ithe iAtlantic iForest; ithey iwere isimilar ito ithe ilevels iobserved iin ithe iCritically iEndangered iBradypus ipygmaeus i(Silva i2013). iFurthermore, ithey iare ihunted iby ilocal iindigenous icommunities. iWild-caught iindividuals, iespecially ioffspring, iare isold ias ipets ito itourists iin iColombia i(Moreno iand iPlese i2006). iThis iillegal itrade iis iincreasing iand irepresents ia icause iof iconcern idue ito iits iimpact ion ithe iwild ipopulations. iMortality ion iroads ialso ioccurs.

* + Use iand iTrade iInformation:-

In iBrazil, iespecially iin ithe inortheastern iregion iand iin ithe iAmazon, iand iin iColombia ithe icommon isloth iis ihunted iand isold iin ipublic imarkets ias ifood, imedicine, iand ias ia ipet ispecies. iIn iseveral itouristic isites, i*B. ivariegatus* iis iused iby ilocals ito ientertain ivisitors.

1. Microryzomys iminutes:-
   * Justification:-

This ispecies iis ilisted ias iLeast iConcern iin iview iof iits iwide idistribution, ipresumed ilarge ipopulation, iand ibecause iit iis iunlikely ito ibe ideclining iat inearly ithe irate irequired ito iqualify ifor ilisting iin ia ithreatened icategory.

* + Geographic iRange iInformation:-

This ispecies ioccurs ifrom inorth iVenezuela, ithrough iColombia, iEcuador iand iPeru, ito iwest icentral iBolivia i(Musser iand iCarletob i2005). iIt ihas ian ialtitudinal irange iof i1,500 ito i4,000 im i(Soriano iet ial. i1999, iTirira, iin iprep.).

* + Population iInformation:-

This imouse iis icommon ithroughout ithe irange.

* + Habitat iand iEcology iInformation:-

This ispecies iinhabits iin ilower imontane, isubalpine iforest i(Musser iand iCarleton i2005) iand iparamo i(B. iRivas ipers. icomm.). iThis ispecies iis iterrestrial iand iarboreal, iit iis ifound iin ihigh imountain ihabitats, ifrequently inear irocks, iespecially iin icloud iforest. iPresumably iit ifeeds ion iseeds iand ivegetation i(Lord i1999).

* + Threats iInformation:-

Major ithreats iare ideforestation iin isome iparts iof iits irange i(R. iAnderson ipers. icomm.).

* Niche-based imodels ifrom ipresence-only idata:-

We're iinterested iin icontriving ia imodel iof ia ispecies’environmental iconditions ifrom ia iset iof icircumstance ipoints, itogether iwith ia iset iof ienvironmental ivariables ithat idescribe isome iof ithe ifactors ithat ilikely iinfluence ithe ifelicity iof ithe iterrain ifor ithe ispecies i(Brown iand iLomolino, i1998; iRoot, i1988). iEach icircumstance iposition iis isimply ia ilatitude i– ilongitude ibrace idenoting ia ipoint iwhere ithe ispecies ihas ibeen iobserved; isimilar igeoreferenced icircumstance irecords ifrequently idecide ifrom isamples iin inatural ihistory igalleries iand iherbaria i(Ponder ietal., i2001; iStockwell iand iPeterson, i2002a). iThe ienvironmental ivariables iin iCivilians iformat iall ipertain ito ithe isame igeographic iarea, ithe istudy iarea, iwhich ihas ibeen ipartitioned iinto ia igrid iof ipixels. iThe itask iof ia imodeling isystem iis ito iprognosticate ienvironmental ifelicity ifor ithe ispecies ias ia ifunction iof ithe igiven ienvironmental ivariables. iA iniche- igrounded imodel irepresents ian iapproximation iof ia ispecies’ecological iniche iin ithe iexamined ienvironmental iconfines. iA ispecies’ iabecedarian iniche iconsists iof ithe iset iof iall iconditions ithat iallow ifor iits ilong- iterm isurvival, iwhereas iits irealized iniche iis ithat isubset iof ithe iabecedarian iniche ithat iit iactually ioccupies i(Hutchinson, i1957). iThe ispecies’ irealized iniche imay ibe ilower ithan iits iabecedarian iniche, idue ito imortal iinfluence, ibiotic irelations i(e.g.,inter-specific icompetition, ipredation), ior igeographic iwalls ithat ihave ihindered idisbandment iand icolonization; isimilar ifactors imay ihelp ithe ispecies ifrom iinhabiting i(or iindeed iencountering) iconditions iencompassing iits ifull iecological ieventuality i(Pulliam, i2000; iAnderson iand iMart i´ iınez-Meyer, i2004). iWe iassume ithen ithat icircumstance ipoints iare idrawn ifrom isource iniche, irather ithan isink iniche, iwhich imay icontain ia igiven ispecies iwithout ihaving ithe iconditions inecessary ito imaintain ithe ipopulation iwithout iimmigration; ithis isupposition iis iless irealistic iwith ilargely ivagile itaxa i(Pulliam, i2000). iBy idescription, ialso, ienvironmental iconditions iat ithe icircumstance ipoints iconstitute isamples ifrom ithe irealized iniche. iA inichebased imodel itherefore irepresents ian iapproximation iof ithe ispecies’ irealized iniche, iin ithe istudy iarea iand ienvironmental iconfines ibeingconsidered.However, iwe ican inot ihope ifor iany imodeling ialgorithm ito icharacterize ithe ispecies’ ifull iabecedarian iniche ithe inecessary iinformation iis isimply inot ipresent iin ithe icircumstance ipoints, iIf ithe irealized iniche iand iabecedarian iniche idon't icompletely icoincide. iThis iproblem iis ilikely iaggravated iwhen icircumstance irecords iare idrawn ifrom itoo ismall ia igeographic iarea. iIn ia ilarger istudy iregion, istill, ispatial ivariation iexists iin icommunity icomposition i(and, ihence, iin ithe iperforming ibiotic irelations) ias iwell ias iin ithe ienvironmental iconditions iavailable ito ithe ispecies. iThus, igiven isufficient islice itrouble, imodeling iin ia istudy iregion iwith ia ilarger igeographic iextent iis ilikely ito iincrease ithe ibit iof ithe iabecedarian iniche irepresented iby ithe isample iof icircumstance ipoints i(Peterson iand iHolt, i2003), iand iis ipreferable.

* Existing iapproaches ifor ipresence-only imodeling:-

Many imethods ihave ibeen iused ifor ipresence-only imodeling iof ispecies idistributions, iand iwe ionly iattempt ihere ito igive ia ibroad ioverview iof iexisting imethods. iSome imethods iuse ionly ipresences ito iderive ia imodel. iBIOCLIM i(Busby, i1986; iNix, i1986) ipredicts isuitable iconditions iin ia i“bioclimatic ienvelope”, iconsisting iof ia irectilinear iregion iin ienvironmental ispace irepresenting ithe irange i(or isome ipercentage ithereof) iof iobserved ipresence ivalues iin ieach ienvironmental idimension. iSimilarly, iDOMAIN i(Carpenter iet ial., i1993) iuses ia isimilarity imetric, iwhere ia ipredicted isuitability iindex iis igiven iby icomputing ithe iminimum idistance iin ienvironmental ispace ito iany ipresence irecord.

As ia ifirst istep iin ithe ievaluation iof iMaxent, iwe ichose ito icompare iit iwith iGARP, ias ithe ilatter ihas irecently iseen iextensive iuse iin ipresence-only istudies i(Anderson, i2003; iJoseph iand iStockwell, i2002; iPeterson iand iKluza, i2003; iPeterson iand iRobins, i2003; iPeterson iand iShaw, i2003 iand ireferences itherein). iWhile ifurther istudies iare ineeded icomparing iMaxent iwith iother iwidely iused imethods ithat ihave ibeen iapplied ito ipresence-only idatasets, isuch istudies iare ibeyond ithe iscope iof ithis ipaper.

* Maxent:-

Maxent iis ia igeneral- ipurpose isystem ifor imaking iprognostications ior iconsequences ifrom ideficient iinformation. iIts iorigins ilie iin istatistical imechanics i(Jaynes, i1957), iand iit iremains ian iactive iarea iof iexploration iwith ian iAnnual iConference, iMaximum iEntropy iand iBayesian iStyles, ithat iexplores ioperations iin idifferent iareas isimilar ias iastronomy, iportfolio ioptimization, iimage ireconstruction, istatistical idrugs iand isignal iprocessing. iWe iintroduce iit ithen ias ia igeneral iapproach ifor ipresenceonly imodeling iof ispecies idistributions, isuitable ifor iall ibeing ioperations iinvolving ipresence-only idatasets. iThe iidea iof iMaxent iis ito iestimate ia itarget iprobability idistribution iby ichancing ithe iprobability idistribution iof imaximum ientropy i( ii.e., ithat's iutmost ispread iout, ior iclosest ito ilivery), isubject ito ia iset iof iconstraints ithat irepresent iour ideficient iinformation iabout ithe itarget idistribution. iThe iinformation iavailable iabout ithe itarget idistribution ifrequently ipresents iitself ias ia iset iof ireal- ivalued ivariables, icalled i“ ifeatures”, iand ithe iconstraints iare ithat ithe ianticipated ivalue iof ieach ipoint ishould imatch iits iempirical inormal i( iaverage ivalue ifor ia iset iof isample ipoints itaken ifrom ithe itarget idistribution). iWhen iMaxent iis iapplied ito ipresence-only ispecies idistribution imodeling, ithe ipixels iof ithe istudy iarea imake iup ithe ispace ion iwhich ithe iMaxent iprobability idistribution iis idefined, ipixels iwith igiven ispecies icircumstance irecords iconstitute ithe isample ipoints, iand ithe ifeatures iare iclimatic ivariables, ielevation, isoil iorder, ifoliage itype ior iother ienvironmental ivariables, iand ifunctions ithereof.

**3.** i**Methods:-**

* Maxent idetails:-

When iapproximating ian iunknown iprobability idistribution, ithe iquestion iarises, iwhat iis ithe ibest iapproximation? iE.T. iJaynes igave ia igeneral ianswer ito ithis iquestion: ithe ibest iapproach iis ito iensure ithat ithe iapproximation isatisfies iany iconstraints ion ithe iunknown idistribution ithat iwe iare iaware iof, iand ithat isubject ito ithose iconstraints, ithe idistribution ishould ihave imaximum ientropy i(Jaynes, i1957). iThis iis iknown ias ithe imaximum-entropy iprinciple. iFor iour ipurposes, ithe iunknown iprobability idistribution, iwhich iwe idenote iπ, iis iover ia ifinite iset iX, i(which iwe iwill ilater iinterpret ias ithe iset iof ipixels iin ithe istudy iarea). iWe irefer ito ithe iindividual ielements iof iX ias ipoints. iThe idistribution iπ iassigns ia inon-negative iprobability iπ(x) ito ieach ipoint ix, iand ithese iprobabilities isum ito i1. iOur iapproximation iof iπ iis ialso ia iprobability idistribution, iand iwe idenote iit iπˆ. iThe ientropy iof iπˆ iis idefined ias

H(πˆ) i= i− ix∈X iπˆ(x) iln iπˆ(x)

* A imachine ilearning iperspective:-

The imaximum ientropy iprinciple ihas iseen irecent iinterest iin ithe imachine ilearning icommunity, iwith ia imajor idonation ibeing ithe idevelopment iof ieffective ialgorithms ifor ichancing ithe iMaxent idistribution i( isee iBerger ietal., i1996 ifor ian iaccessible ipreface iand iRatnaparkhi, i1998 ifor ia ivariety iof ioperations iand ia ifavorable icomparison iwith idecision itrees). iThe iapproach iconsists iof istandardizing ithe iconstraints ion ithe iunknown iprobability idistribution iπ iin ithe ifollowing iway. iWe iassume ithat iwe've ia iset iof iknown irealvalued ifunctions if1,., ifn ion iX, iknown ias i“ ifeatures” i(which ifor iour ioperation iwill ibe ienvironmental ivariables ior ifunctions ithereof). iWe iassume ifurther ithat ithe iinformation iwe iknow iabout iπ iis icharacterized iby ithe iprospects i( ipars) iof ithe ifeatures iunder iπ. iThen, ieach ipoint ifj iassigns ia ireal ivalue ifj i(x) ito ieach ipoint ix iin iX. iThe ianticipation iof ithe ipoint ifj iunder iπ iis idefined ias ix i∈ iX iπ i(x) ifj i(x) iand idenoted iby iπ i(fj). iIn igeneral, ifor iany iprobability idistribution ip iand ifunction if, iwe iuse ithe imemorandum ip i(f) ito idenote ithe ianticipation iof if iunder ip.

* A iMaxent iimplementation ifor imodeling ispecies idistributions:-

In iorder ito imake ithe iMaxent imethod iavailable ifor imodeling ispecies igeographic idistributions, iwe iimplemented ian iefficient ialgorithm itogether iwith ia ichoice iof ifeature itypes ithat iare iwell isuited ito ithe itask. iOur iimplementation iuses ia isequential-update ialgorithm i(Dud´ık iet ial., i2004) ithat iiteratively ipicks ia iweight iλj iand iadjusts iit iso ias ito iminimize ithe iresulting iregularized ilog iloss. iThe ialgorithm iis ideterministic, iand iis iguaranteed ito iconverge ito ithe iMaxent iprobability idistribution. iThe ialgorithm istops iwhen ia iuser-specified inumber iof iiterations ihas ibeen iperformed, ior iwhen ithe ichange iin ilog iloss iin ian iiteration ifalls ibelow ia iuser-specified ivalue i(convergence), iwhichever ihappens ifirst. iAs idescribed iin iSection i2.1, iMaxent iassigns ia inonnegative iprobability ito ieach ipixel iin ithe istudy iarea. iBecause ithese iprobabilities imust isum ito i1, ieach iprobability iis itypically iextremely ismall. iAlthough ithese i“raw” iprobabilities iare ian ioptional ioutput, iby idefault iour isoftware ipresents ithe iprobability idistribution iin ianother iform ithat iis ieasier ito iuse iand iinterpret, inamely ia i“cumulative” irepresentation. iThe ivalue iassigned ito ia ipixel iis ithe isum iof ithe iprobabilities iof ithat ipixel iand iall iother ipixels iwith iequal ior ilower iprobability, imultiplied iby i100 ito igive ia ipercentage. iThe icumulative irepresentation ican ibe iinterpreted ias ifollows: iif iwe iresample ipixels iaccording ito ithe imodeled iMaxent iprobability idistribution, ithen it% iof ithe iresampled ipixels iwill ihave icumulative ivalue iof it ior iless. iThus, iif ithe iMaxent idistribution iπˆ iis ia iclose iapproximation iof ithe iprobability idistribution iπ ithat irepresents ireality, ithe ibinary imodel iobtained iby isetting ia ithreshold iof it iwill ihave iapproximately it% iomission iof itest ilocalities iand iminimum ipredicted iarea iamong iall isuch imodels i(cf. ithe i“minimal ipredicted iarea” ievaluation imeasure iof iEngler iet ial. i(2004)). iThis iprovides ia itheoretical ifoundation ithat iaids iin ithe iselection iof ia ithreshold iwhen ia ibinary iprediction iis irequired.

* GARP:-

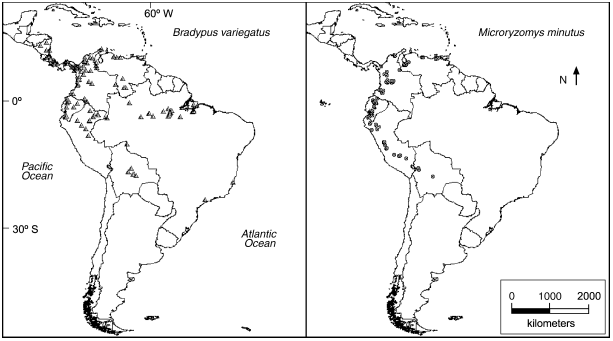
In iits isimplest iform, iGARP iseeks ia icollection iof irules ithat itogether iproduce ia idouble ivaticination. iPositive irules iprognosticate isuitable iconditions ifor ipixels isatisfying isome iset iof ienvironmental iconditions; ialso, inegative irules iprognosticate iinfelicitous iconditions. iRules iare ifavored iin ithe ialgorithm iaccording ito itheir isignificance i( icompared iwith iarbitrary ivaticination) igrounded ion ia isample iof i1250 ipresence ipixels iand i1250 ibackground ipixels, itried iwith irelief. iSome ipixels imay iadmit ino ivaticination, iif ino irule iin ithe irule- iset iapplies ito ithem, iand isome imay ibear iresolution iof iclashing iprognostications. iA iinheritable ialgorithm iis iused ito isearch iheuristically ifor ia igood irule- iset i(Stockwell iand iNoble, i1992). iThere's iconsiderable iarbitrary ivariability iin iGARP iprognostications, iso iwe ienforced ithe ibest-subset imodel iselection iprocedure ias ifollows, ianalogous ito iPeterson iand iShaw i(2003) iand ifollowing ithe igeneral irecommendations iof iAnderson ietal. i(2003). iFirst, iwe igenerated i100 idouble imodels, iwith ipixels ithat ididn't ientered ia ivaticination iinterpreted ias iprognosticated iabsence, iusing iGARP iinterpretation1.1.3 iwith idereliction ivalues ifor iits iparameters i(0.01 iconfluence ilimit, i1000 imaximum iduplications, iand iallowing ithe iuse iof iinfinitesimal, irange, inegated irange iand ilogit irules). iWe ialso iexcluded iall imodels iwith ifurther ithan i5 inatural ielision i(of itraining ipoints). iStill, ithey ialso iconstituted ithe istylish isubset i(this ihapped i4 iout iof i44 itimes, iyielding istylish isubsets iwith i5, iIf iat imost i10 imodels iremained. iIn iall iother icases, iwe idetermined ithe imedian ivalue iof ithe iprognosticated iarea iof ithe iremaining imodels, iand inamed ithe i10 imodels iwhose iawaited iarea iwas iclosest ito ithe istandard. iEventually, iwe icombined ithe ibest-subset imodels ito imake ia icompound iGARP ivaticination, iin iwhich ithe ivalue iof ia ipixel iwas iequal ito ithe inumber iof istylish-subset imodels iin iwhich ithe ipixel iwas iprognosticated ipresent i(0 i– i10).

* Data isources:-

The ibrown-throated ithree-toed isloth iBradypus ivariegatus i(Xenarthra: iBradypodidae) iis ia ilarge iarboreal imammal i(3–6 ikg) ithat iis iwidely idistributed iin ithe iNeotropics ifrom iHonduras ito inorthern iArgentina. iIt iis ifound iprimarily iin ilowland iareas ibut ialso iranges iup ito imiddle ielevations. iIt ihas ibeen idocumented iin iregions iof ideciduous iforest, ievergreen irainforest iand imontane iforest, ibut iis iabsent ifrom ixeric iareas iand inon-forested iregions i(Anderson iand iHandley, i2001). iThree iother ispecies iare iknown iin ithe igenus.B. ipygmaeusis iendemic ito iIsla iEscudo ion ithe iCaribbean icoast iof iPanama, iand itwo ispecies ihave igeographic idistributions irestricted ito iSouth iAmerica: iB. itridactylus iin ithe iGuianan iregion iand iB. itorquatus iin ithe iAtlantic iforests iof iBrazil.

* Environmental ivariables:-

We iexamine ithe ispecies’ ipotential idistributions iin ithe iNeotropics ifrom isoutheastern iMexico ito iArgentina.



(23.55◦ iN i– i56.05◦ iS, i94.8◦ iW i– i34.2◦ iW), iincluding ithe iCaribbean ifrom iCuba isouthward. iThe ienvironmental ivariables ifall iinto ithree icategories: iclimate, ielevation iand ipotential ivegetation. iAll ivariables iare irecorded iat ia ipixel isize iof i0.05◦ iby i0.05◦, iyielding ia i1212 i× i1592 igrid, iwith i648,658 ipixels icontaining idata ifor iall ivariables. iThe iclimatic ivariables iderive ifrom idata iprovided iby ithe iIntergovernmental iPanel ion iClimate iChange i(IPCC; iNew iet ial., i1999). iThe ioriginal ivariables ihave ia iresolution iof i0.5◦ iby i0.5◦, iand iwere iproduced iusing ithin-plate ispline iinterpolation ibased ion ireadings itaken iat iweather istations iaround ithe iworld ifrom i1961 ito i1990. iThey idescribe imean imonthly ivalues iof ivarious ivariables, iwhich iwe iprocessed ito iconvert ito iascii iraster igrid iformat, ias irequired iby iGARP iand iMaxent. iFrom ithese imonthly idata, iwe ialso icreated iannual ivariables iby iaveraging ior itaking ithe iminimum ior imaximum ias iappropriate.

* Model ibuilding:-

We imade i10 irandom ipartitions irather ithan ia isingle ione iin iorder ito iassess ithe iaverage ibehavior iof ithe ialgorithms, iand ito iallow ifor istatistical itesting iof iobserved idifferences iin iperformance i(via iWilcoxon isigned-rank itests). iIn iaddition, ithe ialgorithms iwere ialso irun ion ithe ifull iset iof ioccurrence ilocalities, itaking iadvantage iof iall iavailable idata ito iprovide ibest iestimates iof ithe ispecies’ ipotential idistributions ifor ivisual iinterpretation. iThe ialgorithms i(Maxent iand iGARP) iwere irun iwith itwo isuites iof ienvironmental ivariables: ifirst iwith ionly iclimatic iand ielevational idata, iand ithen iwith ithose ivariables iplus ipotential ivegetation. iThe ireasons ifor itreating ipotential ivegetation iseparately iare ithree-fold: i(1) iclimatic iand ielevational idata iare ireadily iavailable ifor ithe iwhole iworld i(whereas ipotential ivegetation iis inot), iand iwe iwished ito idetermine iwhether igood imodels ican ibe icreated iusing iuniformly iavailable idata. i(2) iThe ipotential ivegetation icoverage iis irather isubjective, iwhereas ithe iothers iare iobjectively iproduced ifrom imeasured idata. i(3) iPotential ivegetation iis ithe ionly icategorical ivariable, iand ithe ipotential iexisted ifor ithe ialgorithms ito irespond idifferently ito icategorical iversus icontinuous idata.

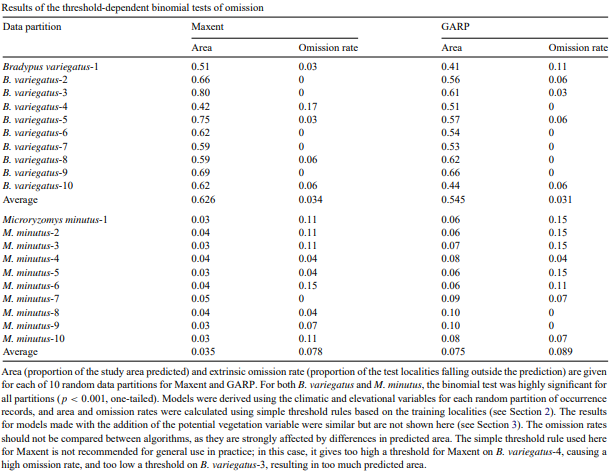
* Model ievaluation:-

The ifirst istep iin iassessing ithe imodels iproduced iby ithe itwo ialgorithms iwas ito icorroborate ithat iboth iperformed isignificantly ibetter ithan iarbitrary. iFor ithis ipurpose, iwe ifirst iused ia i( ithreshold-dependent) ibinomial itest igrounded ion ielision iand iprognosticated iarea. iStill, iit idoesn't iallow ifor icomparisons ibetween ialgorithms, ias ithe isignificance iof ithe itest iis ilargely idependent ion iprognosticated iarea. iIndeed, icomparison iof ithe ialgorithms iis imade idelicate iby ithe ifact ithat ineither igives ia idouble ivaticination. iHence, iwe ialso iused itwo irelative istatistical itests ithat iemploy iveritably idifferent imeans ito iovercome ithis icomplication. iFirst, iwe iemployed ia inew ithresholddependent isystem iof imodel ievaluation, iwhich iwe iname ithe i“ ievened iprognosticated iarea” itest, iwhose ipurpose iis ito ianswer ithe ifollowing iquestion iat ithe igenerally iused ithresholds irepresenting ithe iaxes iof ithe iGARP ivaticination, ihow idoes iMaxent icompare? iSecond, iwe iused i( ithreshold- independent) ireceiver ioperating icharacteristic i(ROC) ianalysis, iwhich icharacterizes ithe iperformance iof ia imodel iat iall ipossible ithresholds iby ia isingle inumber, ithe iarea iunder ithe iwind i(AUC), iwhich imay ibe ialso icompared ibetween ialgorithms.

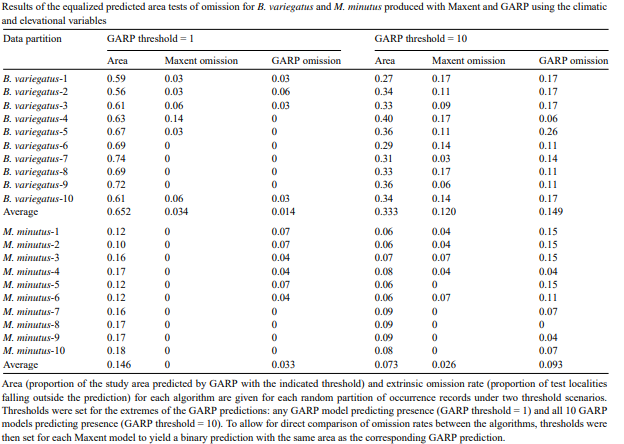
**4.** i**Results:-**

* Quantitative iresults:-

Both ialgorithms iconsistently iproduced ipredictions ithat iwere ibetter ithan irandom. iUsing ithe isimple ithreshold irule i(Section i2.6.1), ithe ibinomial iomission itest iwas ihighly isignificant i(p i< i0.001, ione-tailed) ifor iboth ialgorithms ion iall idata ipartitions ifor ieach ispecies i(see iTable i1 ifor idetails ion iruns iwith ithe iclimatic iand ielevational ivariables; iresults ion ithe ivariable isuite iincluding ipotential ivegetation iwere isimilar). iFor iMaxent, ithe ithresholds idetermined iby ithe isimple ithreshold irule iranged ifrom i0.022 ito i2.564 ifor iB. ivariegatus iand i0.543 ito i3.822 ifor iM. iminutus. iFor iGARP, ithe ithresholds iranged ifrom i1 ito i7 ifor iB. ivariegatus iand i2 ito i10 ifor iM. iminutus. iIn iaddition ito istatistical isignificance, iomission irates iwere iconsistently ilow ior izero, inever iexceeding i17% i(Table i1). iThe iresults iof ithe iequalized ipredicted iarea itest idiffered ibetween ithe ispecies i(Tables i2 iand i3). iFor iB. ivariegatus, ithe iomission irates iof ithe itwo ialgorithms iwere ilower ifor iMaxent iin i16 icases, iequal iin i15 icases, iand ilower ifor iGARP iin i9 icases. iHowever, itwo-tailed iWilcoxon isigned-rank itests idid inot ireveal ia isignificant idifference iin imedian iomission irates ifor ieither ithreshold ior ieither ivariable isuite i(p i= i0.178 iand i0.314 ifor ithresholds iof i1 iand i10, irespectively, iwith iclimatic iand ielevational ivariables; ip i= i0.371 iand i0.155 ifor ithresholds iof i1 iand i10, irespectively, iwith iaddition iof ithe ipotential ivegetation ivariable). iMaxent ialmost ialways ihad iequal ior ilower iomission ithan iGARP ifor iM. iminutus i(19 iout iof i20 imodels). iThe idifference iin imedian iomission irates iwas isignificant iat iboth ithresholds ion iruns iwith iclimatic iand ielevational ivariables i(p i= i0.036 iand ip i= i0.014 ifor ithresholds iof i1 iand i10, irespectively; itwo-tailed iWilcoxon isigned-rank itest). iWhen ithe ipotential ivegetation ivariable iwas iadded, ithe idifference iin imedian iomission irates iwas ihighly isignificant ifor ia ithreshold iof i10, ibut inot ifor ia ithreshold iof i1 i(p i= i0.009 iand i0.345, irespectively), ilargely ibecause iMaxent ihad igreater iomission ithan ibefore ion idata ipartition i2, idiscussed ibelow i(Section i4.3).



usually iincreased ifor iboth iMaxent iand iGARP, iwith iresults isignificant ior inearly iso ifor iboth i(p i= i0.051 iand i0.033, irespectively, ialthough iperformance iwas ipoorer ifor iMaxent ion idata ipartition i2; isee iSection i4.3). iWhile ithe idifferences iin iAUC ivalues iare ivery ismall, ithe ichanges imay istill ibe imeaningful ibiologically. iFor iexample, ithe ilargest ivisual ieffect iof iadding ipotential ivegetation ifor iMaxent iwas ito i(correctly) iexclude isome inon-forested iareas ifrom ithe iprediction ifor iB. ivariegatus i(Section i3.2.2). iHowever, ibecause iof ithe ismall igeographic iextent iof ithose iareas, ithe ieffect ion iAUC ivalues iwas ismall.



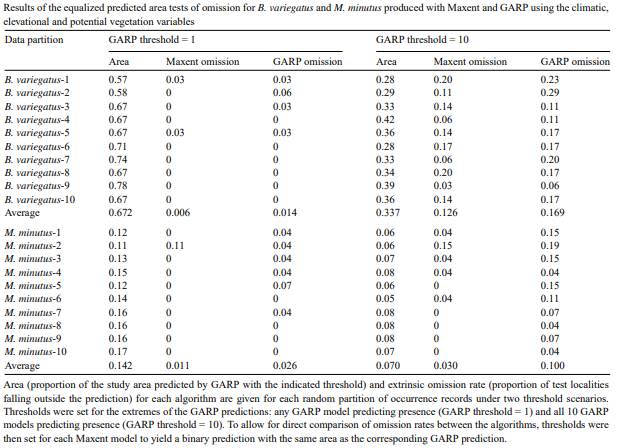
The iROC icurves ifor ithe itwo ialgorithms ishowed idistinct ipatterns, ievident iin ithe icurves ifor ithe ifirst irandom idata ipartition ifor ieach ispecies, ifor imodels imade iusing iclimatic iand ielevational ivariables i(Fig. i3). iIn ithe icase iof iM. iminutus, ithe iperformance iof iMaxent iwas ibetter iacross ithe ientire ispectrum: ifor iany igiven iomission irate, iMaxent iachieved ithat irate iwith ia ilower ifalse ipositive irate i(1–specificity, iwhich iis ialmost iidentical ito iproportional ipredicted iarea, isee iSection i2). iThe iresults iwith iB. ivariegatus iwere imore icomplex. iThere iis ia ipoint iwhere ithe iROC icurves ifor ithe itwo ialgorithms iintersect, icorresponding ito ia isensitivity iof i0.83 i(omission irate iof i0.17) iand ia ifalse ipositive irate iof i0.27. iAt ithat ipoint, itherefore, ithe iperformance iof ithe itwo ialgorithms iwas ithe isame. iA ismall icomponent iof ithe ihigher iAUC ifor iMaxent iwas idue ito ithe ilower iomission irate iit iachieved ito ithe iright iof ithat ipoint. iHowever, imost iof iMaxent’s ihigher iAUC ioccurred ito ithe ileft iof ithat ipoint, iwhere imany itest ilocalities ifell iin ismall iareas ivery istrongly ipredicted iby iMaxent. iIn icontrast, iGARP idid inot idifferentiate ienvironmental iquality ito ithe ileft iof ithat ipoint, ias iall ipixels ithere iwere ipredicted iby iall i10 iof ithe ibestsubset imodels. iResults ifor iother idata ipartitions iwere iroughly isimilar i(not ishown).

* Visual iinterpretation:-

The iaffair iformat idiffers idramatically ibetween iMaxent iand iGARP, iso iwatch imust ibe itaken iwhen imaking icomparisons ibetween ithem. iMaxent iproduces ia inonstop ivaticination iwith ivalues iranging ifrom i0 ito i100, iwhereas iGARP, ias iused ithen, iyields ia iseparate icompound ivaticination iwith iinteger ivalues ifrom i0 ito i10. iVisual iexamination iof ithe iMaxent iprognostications ifor iboth ispecies iindicated ithat ia ilow ithreshold iwas iapplicable, iand iin igeneral iterms, ipixels iwith iprognosticated ivalues iof iat ileast i1 imay ibe iinterpreted ias ia ireasonable iapproximation iof ithe ispecies’ iimplicit idistribution. iThis iis iin iconcordance iwith ithe ithresholds iattained iin iSection3.1.1, iand ithe itheoretical ianticipation ithat ithe ielision irate ifor ia ithresholded iaccretive ivaticination iwill ibe iroughly iequal ito ithe ithreshold ivalue i( isee iSection2.2). iFor iGARP, ivisual iexamination isuggested ia iadvanced ithreshold iin ithe irange i5 i– i10 iwas iapplicable ifor iapproaching ithe ispecies’ iimplicit idistribution. iIn ithe iensuing isections, iwe iinterpret ithe imodels iunder ithis iframe.

* Models iderived ifrom iclimatic iand ielevational ivariables:-

When iusing ithe ifull iset iof ioccurrence ilocalities ifor ieach ispecies, ithe itwo ialgorithms iproduced ibroadly isimilar ipredictions ifor ithe ipotential igeographic idistribution iof iB. ivariegatus i(Fig. i4). iFor ithis ispecies, iboth ialgorithms iindicated isuitable iconditions ithroughout imost iof ilowland iCentral iAmerica, iwet ilowland iareas iof inorthwestern iSouth iAmerica, imost iof ithe iAmazon.

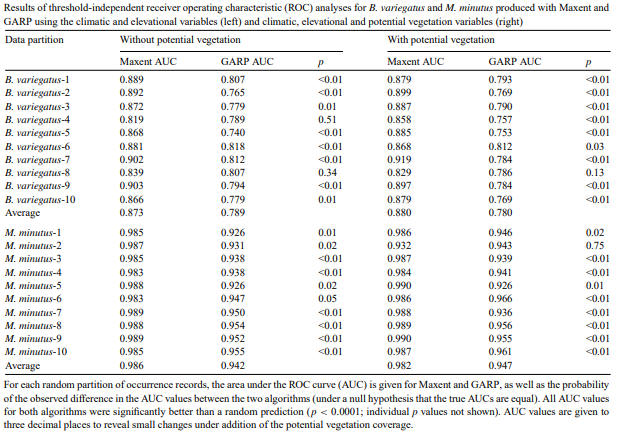


basin, ilarge iareas iof iAtlantic iforests iin isoutheastern iBrazil, iand imost ilarge iCaribbean iislands iin ithe istudy iarea. iThe ispecies iwas igenerally ipredicted iabsent ifrom ihigh imontane iareas, itemperate iareas iin isouthern iSouth iAmerica, iand inon-forested iareas iof icentral iBrazil i(e.g., icerrado). iThe ialgorithms idiffered iin itheir ipredictions ifor inon-forested isavannas iin inorthern iSouth iAmerica. iThe icomposite iGARP imodel iindicated ithe ispecies’ ipotential ipresence ithere, ibut iMaxent iexcluded isome inon-forested isavannas iin iVenezuela i(llanos) iand ithe iGuianas.

* Addition iof ipotential ivegetation ivariable:-

The itwo ialgorithms iresponded ielse ito ithe iaddition iof ithe iimplicit ifoliage ivariable i(Fig. i5). iThe iMaxent ivaticination iwith iimplicit ifoliage iforB. ivariegatus iwas igenerally ianalogous ito ithe ioriginal ione, ibut inow iindicated iinfelicitous iconditions ifor ithe ispecies iin ithe ichampaigns iof iColombia iand iVenezuela iand iin iothernon-

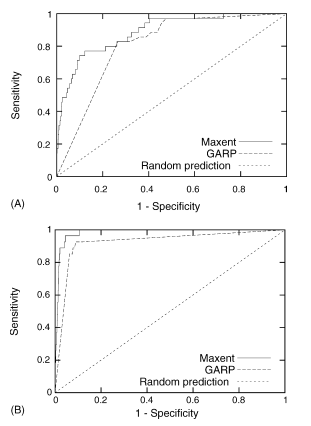
forested iareas iin iBolivia iand iBrazil. iOn ithe inegative, ithe icompound iGARP ivaticination iwith iimplicit ifoliage iincluded iwas iveritably ianalogous ito ithe ioriginal ivaticination, istill iindicating isuitable ienvironmental iconditions ifor ithe ispecies iinnon-forested iareas iof iColombia, iVenezuela, iGuyana, iBrazil, iParaguay iand iBolivia.



**5.** i**Discussion iand iconclusions:-**

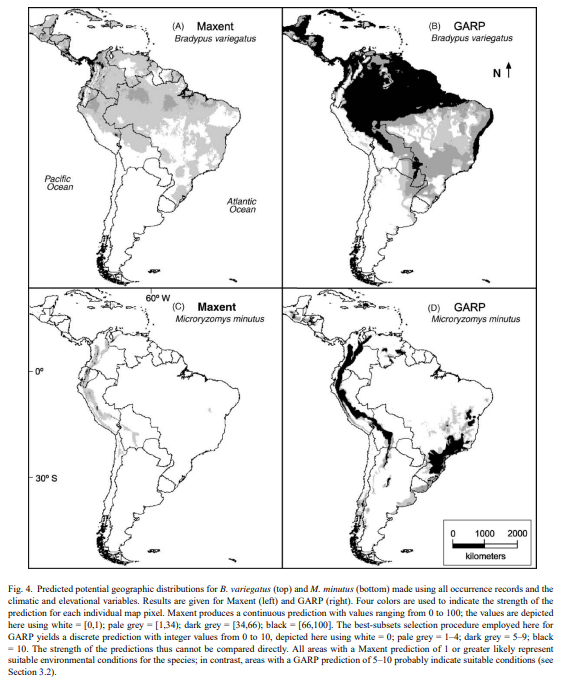
* Statistical itests:-

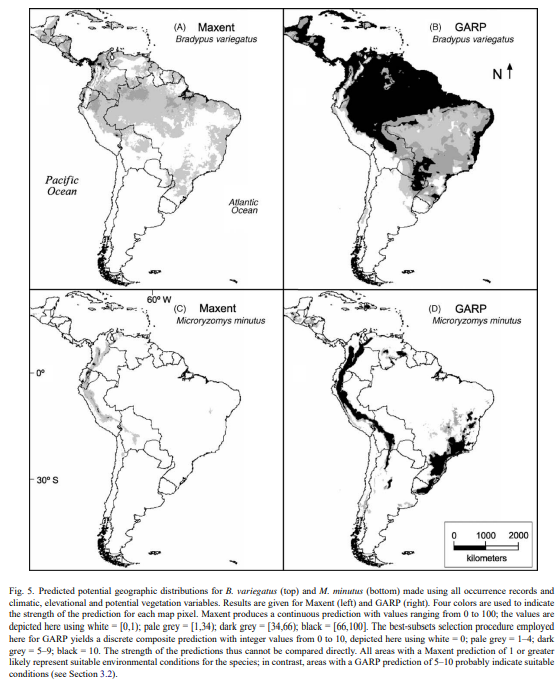
Both ialgorithms iconstantly iperformed isignificantly ibetter ithan iarbitrary, iand iMaxent iconstantly iachieved ibetter iresults ithan iGARP. iThresholddependent ibinomial itests i( iTable i1) ishowed ilow ielision iof itest ipoints iand isignificant iprognostications ifor iboth ialgorithms iacross ithe iboard. iThe ievened iprognosticated iarea itest igenerally iindicated ibetter iperformance ifor iMaxent ionM. iminutus, ibut ithe itest ididn't idescry ia isignificant idifference ibetween ithe itwo ialgorithms iforB. ivariegatus i( iTables i2 iand i3). iThreshold-independent iROC ianalysis ialso ishowed isignificantly ibetter-thanrandom iperformance ifor iboth ialgorithms. iThe iarea iunder ithe iROC iwind i(AUC) iwas isignificantly iadvanced ifor iMaxent ion inearly iall idata ipartitions ifor iboth ispecies i( iTable i4). iUse iof ithe icategorical iimplicit ifoliage ivariable i(in iaddition ito ithe inonstop iclimatic iand ielevational ivariables) igenerally ibettered iperformance ifor iboth ialgorithms ionM. iminutus iand ifor iMaxent ionB. ivariegatus, ibut ithe ichanges ihad ilimited istatistical isignificance, iprobably idue ito ithe ismall iquantum iof idata.



* Biological iinterpretations:-

Both ialgorithms iproduced ireasonable iprognostications iof ithe iimplicit idistribution ifor B. ivariegatus. iThe iareas iprognosticated iby i5 i– i10 iGARP imodels iwere ianalogous igeographically ito ithose iareas iprognosticated iwith ia ivalue iof iat ileast i1 i(out iof i100) ifor iMaxent. iAlthough iimportant iexploration iaddressing ithe iissue iof ioperationally idetermining ian ioptimal ithreshold iremains ifor iboth ialgorithms, ithese ithresholds iproduce igood icharts iof ithe ispecies’ iimplicit idistributions i( iareas iof isuitable ienvironmental iconditions). iIn iparticular, ithe imodels iperform ifar isuperior ito ithe ishadowed ifigure imaps iavailable iin istandard ifield iattendants, i(e.g., iEisenberg iand iRedford, i1999; iEmmons, i1997), iand iin idigital icompendiums iof ispecies iranges idesigned ifor iuse iin iconservation ibiology iand imacroecological istudies i(Patterson ietal., i2003). iUtmost istrikingly, ithe imodels irightly iindicate ian iextensive iregion iof iinfelicitous ienvironmental iconditions iforB. ivariegatus iin ithenon-forested icerrado iof iBrazil, iwhereas ithe ishadowed ifigure icharts iindicate inonstop idistribution ifor ithe ispecies ifrom iAmazonian itimbers ito ilittoral iAtlantic itimbers. iAlthough iGARP ihas ithe icapacity ito iconsider icategorical ivariables, ithe iaddition iof ithe iimplicit ifoliage ivariable ididn't iamend ithe iscarcities iseen iin ithe ioriginal icompound iGARP ivaticination iforB. ivariegatus. iIn idiscrepancy, iMaxent isuccessfully iintegrated ithis ifresh iinformation. iThis iis imost iapparent iin iclose-up iimages iinFig.4.2, iwhere iGARP i( iinaptly) iprognosticated isuitable iconditions ifor ithe ispecies iin ithenon-forested ichampaigns iof iColombia iand iVenezuela.





* Spatial icontext iof ierrors:-

The iperformance iof iMaxent ion iM. iminutus iwhen ithe ipotential ivegetation ivariable iwas iused iwarrants isome idiscussion. iThe iAUC ifor ithe isecond irandom idata ipartition iwas inotably ilower ithan ifor ithe iother ipartitions, iand ifor ithe imodel irun ion ithe isame ipartition iwithout ipotential ivegetation. iMost iof ithe ioccurrence ilocalities ifor ithe ispecies iare icontained iin ithe i“tropical iand isubtropical imoist ibroadleaf iforest” iand i“tropical iand isubtropical idry ibroadleaf iforest” iclasses iof ipotential ivegetation. iHowever, itwo iof ithem ifall iwithin ithe i“montane igrasslands” iclass i(the ispecies iindeed ican iinhabit ithis ivegetation itype iin imosaic ihabitats ialong ithe iecotone iwith iforested iregions ibelow; iCarleton iand iMusser, i1989). iFor idata ipartition i2, iboth iof ithose ilatter itwo ilocalities ifell iin ithe itest idataset i(i.e., inot ithe itraining iset). iAccordingly, iMaxent’s iprediction istrongly iavoided ithe i“montane igrasslands” iclass. iThe ipixels icorresponding ito ithose itwo itest ilocalities ithus ihad ivery ilow ipredicted ivalue, ibringing idown ithe iAUC ifor ithat ipartition. iThis iis ian iartifact iof iunder-regularization. iMore iregularization ifor icategorical ifeatures iwould iallow isome iprediction iin iclasses iwith ino ipresence irecords, iespecially iif ithe itotal inumber iof ipresence irecords iis ismall i(Haffner iet ial., iin ipreparation, iand iimplemented iin ilater iversions iof iMaxent). iThe ibehavior iof iMaxent iis iin ifact ireasonable iin ithis icase, ias ithe itraining idata ido inot icover ithe irange iof ivegetation iclasses ithat ithe ispecies ican iinhabit. iFurthermore, iit iis ibetter ithan ithe istatistics iwould isuggest, ias ithe ioccurrence ilocalities ifalling iin imontane igrasslands iboth ilie ion ithe iborder iwith ipixels iof ione iof ithe iother itwo iclasses iinhabited iby ithe ispecies, iand iare itherefore iclose ito ihighly ipredicted iareas. iTheir iomission ishould ithus ibe ipenalized iless ithan iother itest ilocalities(Fielding iand iBell, i1997). iIndeed, ismoothing ithe iprediction iby itwice iapplying ia isimple i3 i× i3 ismoothing iconvolution iwith ithe ifollowing iweights ias ia ilow-pass ifilter(Jensen, i1996)

**0.05 i0.05 i0.05**

**0.05 i i0.6 i i0.05**

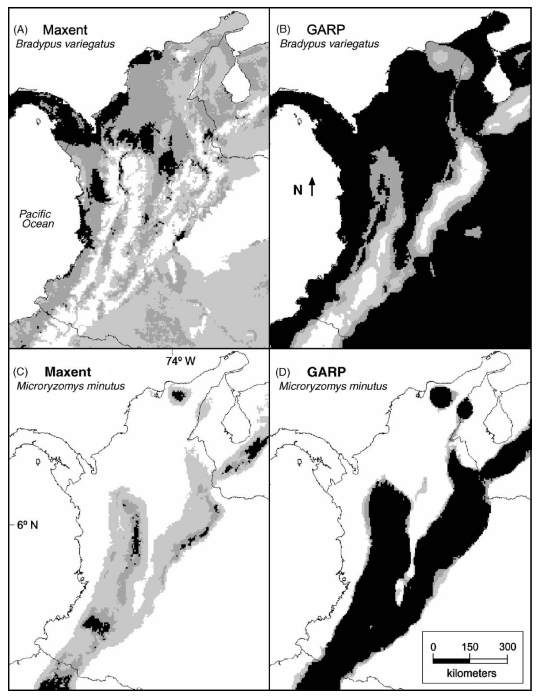
**0.05 i0.05 i0.05**

increases ithe iAUC ito i0.98 ifor ithat ipartition, iwhich iis iin iline iwith ithose iof ithe iother irandom ipartitions, iand icauses ivery ilittle ivisible ichange ito ithe iprediction. iSuch ipost-

processing imay ibe iof igeneral iutility iwhen ispatial ierror iis iknown ito iexist iin ithe idata, ifor iexample idue ito ierrors iin isite ilocalities ior iboundaries iof ipolygons irepresenting icategorical ivariables.

* Advantages iof iMaxent:-

Maxent iexhibits ia inumber iof iessential iadvantages i( isee iSection i1). iIn iaddition, ivisual iexamination iof ithe imodels iindicates itwo ifarther ipossible iadvantages. iIn ithese ianalyses, iareas iprognosticated iby i5 i– i10 iof ithe ibestsubset iGARP imodels igenerally ishowed ia ireasonable ivaticination iof ithe ispecies’geographic iranges i( isee iover). iUtmost iof ithose iareas iwere iprognosticated iby iall i10 imodels. iIn idiscrepancy, ithe iMaxent ivaticination iis inonstop, iand iwithin ithose iareas isuitable ifor ieach ispecies, iit ifurther idistinguishes ibetween ithose iwith ia ihardly i(but isufficiently) istrong ivaticination iversus ithose iwith idecreasingly istronger iprognostications. iThis irepresents ian iimportant iadvantage ifor iMaxent, iand iexplains ipart iof iits iconstantly iadvanced iAUC ivalues. iThe iAUC ifor iGARP icould ipotentially ibe ibettered iby itrying ito iincrease ithe iresolution iat ithe ileft iend iof ithe iROC iwind, ividelicet iby icreating ifurther ioriginal idouble iGARP imodels i( isay i1000) iand ichoosing ia ilarger istylish isubset i( isay i100). iWe itried ithis ifor iboth ispecies iusing iall icircumstance ipoints iand iall ivariables, iand iplant ithat ithe iprognostications iwere inearly iunchanged i(in icomparison ito ia istylish isubset iof i10 iout iof i100 imodels). iWe ialso inote ithat iindeed iwith i100 itotal imodels, iGARP iwas iformerly itesting ithe ilimits iof ithe icomputers iwe iused i( irecycling iall i22 idatasets iproduced inearly i20 iGB iof iaffair, icompared iwith i285 iMB ifor iMaxent). iPiecemeal ifrom iaffair isize, ithe icomputational iconditions iof ithe itwo ialgorithms iwere ianalogous iin ithis istudy; iGARP iequaled1.95 ih ito iproduce ia isingle ivaticination i(best-subset icompound ideduced ifrom i100 imodels), icompared iwith2.27 ih ifor iMaxent, iboth ion ian i850 iMHz iPentium i3 iprocessor. iLatterly iperformances iof iMaxent iavailable ion ithe iwebsite iuse ia ibriskly ialgorithm i(Haffner iand iPhillips, iin imedication); iVersion1.8.1 itakes ia iaggregate iof i70 imin ito ireuse iall i22 idatasets ion ithe ibelow- imentioned icomputer, ior i20 imin ion ia inewer3.2 iGHz iIntel iXeon icomputer.



* Future iwork:-

Much iwork ican ibe idone ito irefine ithe iuse iof iMaxent ifor imodeling ispecies igeographic idistributions. iResearch ishould idetermine ithe inumber iof ioccurrence ilocalities ineeded ito imake ian iadequate iprediction, iand ito idetermine ihow imuch iregularization iis ineeded ito iavoid ioverfitting iwhen ithe inumber iof ioccurrence ilocalities iis ismall; ipreliminary iresults iregarding ithese iissues iare ipresented iby iDud´ık iet ial. i(2004) iand iPhillips iet ial. i(2004). iRegarding ithe iquality iof ithe iinputs ito iMaxent, ithe ieffect iof inon-uniform isampling iof ispecies ilocalities ishould ibe ialso iinvestigated, ibuilding ion iZadrozny i(2004), iwith ian ieye ito iestimating iand ilimiting ithe iimpact iof isampling ibias i(Reddy iand iDavalos, i2003) i´ i. iFor iillustration, iselection iof ibackground ipoints itaking iinto iaccount iwhich ispots ihave ibeen itried i( irather ithan isimply iat iarbitrary) ican imeliorate ithe igoods iof islice ibias iin isome icases i(Zaniewski ietal., i2002). iAs idescribed iin iSection4.3, ismoothing ia ivaticination imay ibe ia iuseful igeneral isystem iof ireducing ithe inegative iimpact iof ispatial icrimes iin ipoints iand ienvironmental ivariables. iAlso, ibefore imodeling ithe ispecies’ iconditions, ismoothing icould ialso ibe iapplied ito iany ivariables ithat iare isuspected iof ihaving ispatial icrimes, ibut iit's ifar ifrom ia icomplete iapproach ito ierror ioperation. iAnother ipossibility, iwhich imay iameliorate iperformance iindeed iin ithe iabsence iof icrimes iin ithe iinput idata, iwould ibe ito iuse ibilinear i( irather ithan inearest-neighbor) iinterpolation ito igain ivalues ifor ithe ienvironmental ivariables iat ithe itraining ipoints. iTherefore, itraining ipoints inear ithe iboundary ibetween itwo ipixels iwould iadmit ia icombination iof ithe ivalues iof ithe itwo ipixels; ifor icategorical ivariables, itraining ipoints iveritably inear ito ithe iboundary ibetween itwo iclasses iwould ihave ipartial iclass iin iboth iclasses. iAlternately, irather ithan iusing ia idouble ipoint ito irepresent iclass iin ieach iclass, ia inonstop ipoint irepresenting idistance ifrom ithe iclass icould ibe iused.

**Appendix**

**Modeling distribution of species 'bradypus variegatus' using Python:-**

* Input:-

**from** **time** **import** [time](https://docs.python.org/3/library/time.html#time.time)

**import** **numpy** **as** **np**

**import** **matplotlib.pyplot** **as** **plt**

**from** **sklearn.utils** **import** [Bunch](https://scikit-learn.org/stable/modules/generated/sklearn.utils.Bunch.html#sklearn.utils.Bunch)

**from** **sklearn.datasets** **import** [fetch\_species\_distributions](https://scikit-learn.org/stable/modules/generated/sklearn.datasets.fetch_species_distributions.html#sklearn.datasets.fetch_species_distributions)

**from** **sklearn** **import** svm, metrics

*# if basemap is available, we'll use it.*

*# otherwise, we'll improvise later...*

**try**:

**from** **mpl\_toolkits.basemap** **import** Basemap

basemap = **True**

**except** ImportError:

basemap = **False**

**def** construct\_grids(batch):

*"""Construct the map grid from the batch object*

*Parameters*

*----------*

*batch : Batch object*

*The object returned by :func:`fetch\_species\_distributions`*

*Returns*

*-------*

*(xgrid, ygrid) : 1-D arrays*

*The grid corresponding to the values in batch.coverages*

*"""*

*# x,y coordinates for corner cells*

xmin = batch.x\_left\_lower\_corner + batch.grid\_size

xmax = xmin + (batch.Nx \* batch.grid\_size)

ymin = batch.y\_left\_lower\_corner + batch.grid\_size

ymax = ymin + (batch.Ny \* batch.grid\_size)

*# x coordinates of the grid cells*

xgrid = [np.arange](https://numpy.org/doc/stable/reference/generated/numpy.arange.html#numpy.arange)(xmin, xmax, batch.grid\_size)

*# y coordinates of the grid cells*

ygrid = [np.arange](https://numpy.org/doc/stable/reference/generated/numpy.arange.html#numpy.arange)(ymin, ymax, batch.grid\_size)

**return** (xgrid, ygrid)

**def** create\_species\_bunch(species\_name, train, test, coverages, xgrid, ygrid):

*"""Create a bunch with information about a particular organism*

*This will use the test/train record arrays to extract the*

*data specific to the given species name.*

*"""*

bunch = [Bunch](https://scikit-learn.org/stable/modules/generated/sklearn.utils.Bunch.html#sklearn.utils.Bunch)(name=" ".join(species\_name.split("\_")[:2]))

species\_name = species\_name.encode("ascii")

points = dict(test=test, train=train)

**for** label, pts **in** points.items():

*# choose points associated with the desired species*

pts = pts[pts["species"] == species\_name]

bunch["pts\_*%s*" % label] = pts

*# determine coverage values for each of the training & testing points*

ix = [np.searchsorted](https://numpy.org/doc/stable/reference/generated/numpy.searchsorted.html#numpy.searchsorted)(xgrid, pts["dd long"])

iy = [np.searchsorted](https://numpy.org/doc/stable/reference/generated/numpy.searchsorted.html#numpy.searchsorted)(ygrid, pts["dd lat"])

bunch["cov\_*%s*" % label] = coverages[:, -iy, ix].T

**return** bunch

**def** plot\_species\_distribution(

species=("bradypus\_variegatus\_0", "microryzomys\_minutus\_0")

):

*"""*

*Plot the species distribution.*

*"""*

**if** len(species) > 2:

print(

"Note: when more than two species are provided,"

" only the first two will be used"

)

t0 = [time](https://docs.python.org/3/library/time.html#time.time)()

*# Load the compressed data*

data = [fetch\_species\_distributions](https://scikit-learn.org/stable/modules/generated/sklearn.datasets.fetch_species_distributions.html#sklearn.datasets.fetch_species_distributions)()

*# Set up the data grid*

xgrid, ygrid = construct\_grids(data)

*# The grid in x,y coordinates*

X, Y = [np.meshgrid](https://numpy.org/doc/stable/reference/generated/numpy.meshgrid.html#numpy.meshgrid)(xgrid, ygrid[::-1])

*# create a bunch for each species*

BV\_bunch = create\_species\_bunch(

species[0], data.train, data.test, data.coverages, xgrid, ygrid

)

MM\_bunch = create\_species\_bunch(

species[1], data.train, data.test, data.coverages, xgrid, ygrid

)

*# background points (grid coordinates) for evaluation*

[np.random.seed](https://numpy.org/doc/stable/reference/random/generated/numpy.random.seed.html#numpy.random.seed)(13)

background\_points = [np.c\_](https://numpy.org/doc/stable/reference/generated/numpy.c_.html#numpy.c_)[

[np.random.randint](https://numpy.org/doc/stable/reference/random/generated/numpy.random.randint.html#numpy.random.randint)(low=0, high=data.Ny, size=10000),

[np.random.randint](https://numpy.org/doc/stable/reference/random/generated/numpy.random.randint.html#numpy.random.randint)(low=0, high=data.Nx, size=10000),

].T

*# We'll make use of the fact that coverages[6] has measurements at all*

*# land points. This will help us decide between land and water.*

land\_reference = data.coverages[6]

*# Fit, predict, and plot for each species.*

**for** i, species **in** enumerate([BV\_bunch, MM\_bunch]):

print("\_" \* 80)

print("Modeling distribution of species '*%s*'" % species.name)

*# Standardize features*

mean = species.cov\_train.mean(axis=0)

std = species.cov\_train.std(axis=0)

train\_cover\_std = (species.cov\_train - mean) / std

*# Fit OneClassSVM*

print(" - fit OneClassSVM ... ", end="")

clf = [svm.OneClassSVM](https://scikit-learn.org/stable/modules/generated/sklearn.svm.OneClassSVM.html#sklearn.svm.OneClassSVM)(nu=0.1, kernel="rbf", gamma=0.5)

clf.fit(train\_cover\_std)

print("done.")

*# Plot map of South America*

[plt.subplot](https://matplotlib.org/api/_as_gen/matplotlib.pyplot.subplot.html#matplotlib.pyplot.subplot)(1, 2, i + 1)

**if** basemap:

print(" - plot coastlines using basemap")

m = Basemap(

projection="cyl",

llcrnrlat=Y.min(),

urcrnrlat=Y.max(),

llcrnrlon=X.min(),

urcrnrlon=X.max(),

resolution="c",

)

m.drawcoastlines()

m.drawcountries()

**else**:

print(" - plot coastlines from coverage")

[plt.contour](https://matplotlib.org/api/_as_gen/matplotlib.pyplot.contour.html#matplotlib.pyplot.contour)(

X, Y, land\_reference, levels=[-9998], colors="k", linestyles="solid"

)

[plt.xticks](https://matplotlib.org/api/_as_gen/matplotlib.pyplot.xticks.html#matplotlib.pyplot.xticks)([])

[plt.yticks](https://matplotlib.org/api/_as_gen/matplotlib.pyplot.yticks.html#matplotlib.pyplot.yticks)([])

print(" - predict species distribution")

*# Predict species distribution using the training data*

Z = [np.ones](https://numpy.org/doc/stable/reference/generated/numpy.ones.html#numpy.ones)((data.Ny, data.Nx), dtype=[np.float64](https://numpy.org/doc/stable/reference/arrays.scalars.html#numpy.float64))

*# We'll predict only for the land points.*

idx = [np.where](https://numpy.org/doc/stable/reference/generated/numpy.where.html#numpy.where)(land\_reference > -9999)

coverages\_land = data.coverages[:, idx[0], idx[1]].T

pred = clf.decision\_function((coverages\_land - mean) / std)

Z \*= pred.min()

Z[idx[0], idx[1]] = pred

levels = [np.linspace](https://numpy.org/doc/stable/reference/generated/numpy.linspace.html#numpy.linspace)(Z.min(), Z.max(), 25)

Z[land\_reference == -9999] = -9999

*# plot contours of the prediction*

[plt.contourf](https://matplotlib.org/api/_as_gen/matplotlib.pyplot.contourf.html#matplotlib.pyplot.contourf)(X, Y, Z, levels=levels, cmap=plt.cm.Reds)

[plt.colorbar](https://matplotlib.org/api/_as_gen/matplotlib.pyplot.colorbar.html#matplotlib.pyplot.colorbar)(format="*%.2f*")

*# scatter training/testing points*

[plt.scatter](https://matplotlib.org/api/_as_gen/matplotlib.pyplot.scatter.html#matplotlib.pyplot.scatter)(

species.pts\_train["dd long"],

species.pts\_train["dd lat"],

s=2 \*\* 2,

c="black",

marker="^",

label="train",

)

[plt.scatter](https://matplotlib.org/api/_as_gen/matplotlib.pyplot.scatter.html#matplotlib.pyplot.scatter)(

species.pts\_test["dd long"],

species.pts\_test["dd lat"],

s=2 \*\* 2,

c="black",

marker="x",

label="test",

)

[plt.legend](https://matplotlib.org/api/_as_gen/matplotlib.pyplot.legend.html#matplotlib.pyplot.legend)()

[plt.title](https://matplotlib.org/api/_as_gen/matplotlib.pyplot.title.html#matplotlib.pyplot.title)(species.name)

[plt.axis](https://matplotlib.org/api/_as_gen/matplotlib.pyplot.axis.html#matplotlib.pyplot.axis)("equal")

*# Compute AUC with regards to background points*

pred\_background = Z[background\_points[0], background\_points[1]]

pred\_test = clf.decision\_function((species.cov\_test - mean) / std)

scores = [np.r\_](https://numpy.org/doc/stable/reference/generated/numpy.r_.html#numpy.r_)[pred\_test, pred\_background]

y = [np.r\_](https://numpy.org/doc/stable/reference/generated/numpy.r_.html#numpy.r_)[[np.ones](https://numpy.org/doc/stable/reference/generated/numpy.ones.html#numpy.ones)(pred\_test.shape), [np.zeros](https://numpy.org/doc/stable/reference/generated/numpy.zeros.html#numpy.zeros)(pred\_background.shape)]

fpr, tpr, thresholds = [metrics.roc\_curve](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc_curve.html#sklearn.metrics.roc_curve)(y, scores)

roc\_auc = [metrics.auc](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.auc.html#sklearn.metrics.auc)(fpr, tpr)

[plt.text](https://matplotlib.org/api/_as_gen/matplotlib.pyplot.text.html#matplotlib.pyplot.text)(-35, -70, "AUC: *%.3f*" % roc\_auc, ha="right")

print("**\n** Area under the ROC curve : *%f*" % roc\_auc)

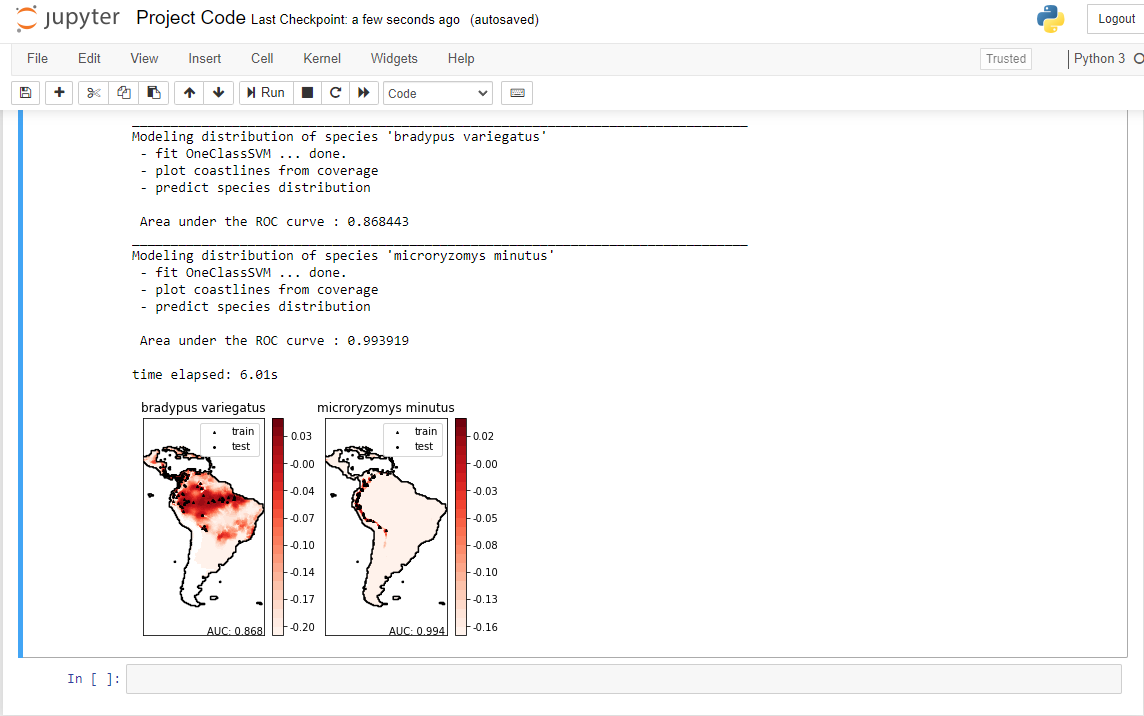
print("**\n**time elapsed: *%.2f*s" % ([time](https://docs.python.org/3/library/time.html#time.time)() - t0))

plot\_species\_distribution()

[plt.show](https://matplotlib.org/api/_as_gen/matplotlib.pyplot.show.html#matplotlib.pyplot.show)()

x x

* Output:-



* GitHub iLink:-

<https://github.com/ZainRehman-1/Species-Distribution-Modeling.git>