TELKOMNIKA, Vol.16, No.2, April 2018, pp. 703~712
ISSN: 1693-6930, accredited **A** by DIKTI, Decree No: 58/DIKTI/Kep/2013
DOI: 10.12928/TELKOMNIKA.v16i2.6894

Predicting the Spread of *Acacia Nilotica* Using Maximum Entropy Modeling

Budi Arif Dermawan*[1], Yeni Herdiyeni[2], Lilik Budi Prasetyo[3], Agung Siswoyo[4]
[1]Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang, Jl. H.S. Ronggowaluyo Telukjambe Timur, Karawang 41361, Indonesia
[2]Department of Computer Science, Bogor Agricultural University, Jl. Raya Dramaga, Kampus Dramaga, Bogor 16680, Indonesia
[3]Department of Forest Resources Conservation & Ecotourism, Bogor Agricultural University, Jl. Raya Dramaga, Kampus Dramaga, Bogor 16680, Indonesia
[4]The Ministry of Environment and Forestry, Bromo Tengger Semeru National Park, East Java, Indonesia
*Corresponding author, e-mail: budi.arif@staff.unsika.ac.id

Abstract

Acacia nilotica planted in Baluran National Park aims to prevent the spread of fire from savanna to teak forest became developed into invasive and led to a decrease in the quality and quantity of savannas. Therefore, it is required to predict the spread of A. nilotica to minimize the impacts of invasion on savanna area. The study aims to identify environmental factors which affect spread of A. nilotica. Furthermore, the spread of A. nilotica is predicted using Maximum Entropy. Maximum Entropy is efficient model since it uses presence-only data while the most of other models use presence and absence data. The experimental results reveal six environmental factors, including elevation, slope, NDMI, NDVI, distance from the river, and temperature were identified affecting the spread of A. nilotica. The most dominant environmental factors were elevation and temperature with 40% and 39.6% contributions. Maximum Entropy performed well in predicting the spread of A. nilotica, it was indicated by AUC value of 0.938.

Keywords*: acacia nilotica, baluran national park, invasive species, maximum entropy, prediction*

1. Introduction

Indonesia has the second highest biodiversity in the world after Brazil because there are many ecosystems in it. One of them is the savanna ecosystems in Baluran National Park which is known as the replica of savanna in Africa. The savanna ecosystem soil is formed by fires occurred repeatedly. A disturbance in the Baluran National Park savanna preservation will influence other various ecosystems [1]. The disruptions are generally in the form of cattle grazing and agricultural land use [2].

Another major threat comes from the extension of the uncontrolled invasion of the *A. nilotica* species that was originally imported for the purpose of isolating vegetation from fires [1]. The larger spread and invasion of the species are the significant indicator of the disturbance and threat to the ecosystem. The growth and spread rate of this exotic species can reduce both quality and quantity of the savanna in Baluran National Park. These decreases effect the changes of herbivorous wildlife behavior which originally consumed grass and turn out to be consuming seeds of *A. nilotica*. This happens because there is a friction in the existence of the main feed by *A. nilotica* species. There is also problem arise from the area where the invasion of *A. nilotica* replaced by that of another invasive species like *Ocimum basilicum* and *Thespesia lampas*, which are less popular for animals [3].

India, Pakistan, and Africa are believed to be the origin of *A. nilotica* species [4]. *A. nilotica* found commonly on the soil containing high clay. It can also grow in the deep sandy loam and also in an area whose high rainfall [5]. *A. nilotica* which was introduced to Indonesia is an indica sub species. *A. nilotica* was intended to be a commercial high quality gum producer. The introduction did not reach the target since the gum production was still low. The introduction then continued to the Baluran National Park as a fire barrier to avoid the spread of fire from

Received July 27, 2017; Revised February 12, 2018; Accepted March 2, 2018

savanna to the forest [6]. Nevertheless, the invasion of *A. nilotica* caused the distress of other various types of vegetation as the main element of the Baluran grasslands.

A. nilotica is a plant species whose many benefits. However, it has not been explored optimally in Indonesia. In the other hand, the presence of *A. nilotica* also causes problems as it is an invasive plant [7]. Therefore, a prediction system using Species Distribution Models is required to prevent its spread by understanding the types of habitat of *A. nilotica*. Species Distribution Models (SDMs) offers a quantitative approach to identify the relations between species and environment together with how these relations influence the geographic distributions of the species (species distribution) [8].

Some approaches used commonly for Species Distribution Modeling are linear regression [9], multivariate (Principal Component Analysis) [10-12], GLM [13-14], and Maxent [15]. The basis of SDM is located on the use of presence and absence data obtained from observation, specimens' records and literature. The problem in obtaining absence data occurs during the observation. Moreover, after obtaining the absence data, the data was not reliable because the amount was limited. According to [16] in his observation, it showed that the use of Maxent model resulted stable and reliable prediction and surpass some other methods in the use of presence-only data.

The Maxent model is chosen because it offers several advantages such as: (1) it can use input presence-only data; (2) it can use input variable in the form of either continuous and categorical data; (3) it produces a stable and reliable prediction accuracy when the data condition is not complete and there is a few sample size; (4) it can directly produce a map of the suitability of spatial habitat explicitly; (5) and encompasses a jackknife test feature which can be used to evaluate environmental variable which is considered to be important [16].

Some authors have used the maximum entropy modeling approach for species distribution. [17] used the maximum entropy model to identify environmental factors affecting the presence of the Sumatran Rhinos. [16] used the maximum entropy model to predict climate change quantitatively on Riparian species. [18] used the presence and absence data with logistic regression method to analyze the suitability of habitat for *A. nilotica*. In addition, the maximum entropy method can also be used in the field of electronics. [19] used the principle of maximum entropy to determine the prior distribution in the evaluation of the reliability of a low voltage switch.

This study aims to identify environmental factors that have an important influence on the existence of *A. nilotica.* The Maximum Entropy Modeling is relevant to be used as a predictive modeling factor. The data used in this study include presence-only and environmental data as the development of the research by [18]. The data and information play an important role in determining habitat suitability of *A. nilotica* and consequently, the prevention efforts toward the expansion of the *A. nilotica* can be encouraged. In addition, the information obtained can be used to determine a new potential extension location of the invasion in the area.

2. Research Method

Figure 1 shows the method of this research. To predict the spread of *A. nilotica* invasive species in Baluran National park, there are several steps: data collection, preprocessing, overlaying environmental data, multicollinearity test, data partition, maximum entropy modeling, calibration and evaluation. The tools used to build the prediction model in this research are R software [20] to build the model and Quantum GIS for spatial analysis.

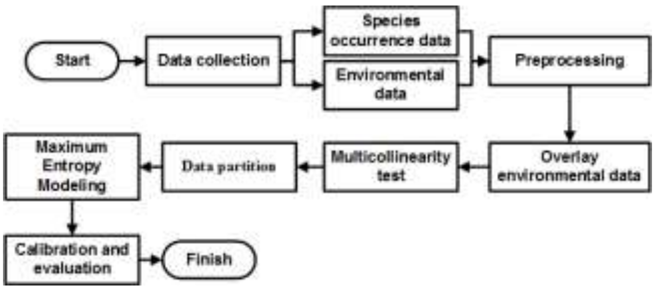

Figure 1. Research method

2.1. Data Collection

The data used in this research are observation data from research [18] in Baluran National Park Situbondo, East Java. The observational data consisted of *Acacia nilotica* distribution data and environmental data affecting the spread of *A. nilotica*. The distribution data used in this research is presence data of *A. nilotica*, while environmental data consist of elevation, slope, NDMI, NDVI, distance from the river, and temperature shown in Table 1.

Table 1. Environment Variable

No	Variables	Units	Data scale	Data source	Extraction technique
1	Elevation	Meter	Ratio	Aster G DEM (http://earthexplorer.usgs.gov/)	Spatial analysis
2	NDMI	-	Ordinal	Landsat satellite imagery 8 OLI (row/path: 121/065)	$\frac{(NIR-IR)}{(NIR+IR)}$
3	NDVI	-	Ordinal	Landsat satellite imagery 8 OLI (row/path: 121/065)	$\frac{(NIR-Red)}{(NIR+Red)}$
4	Distance from the river	Meter	Ratio	Map distance from the river (RBI)	Spatial analysis with *Euclidean Distance*
5	Slope	Percent (%)	Ratio	Aster G DEM (http://earthexplorer.usgs.gov/)	Analysis of the slope of the surface topography
6	Temperature	Degree (^{o})	Ratio	Surface temperature map	Spatial analysis

2.2. Preprocessing

The presence data is organized in comma separated value (CSV). The spatial data of environmental variable must be in raster format, the same extent, and geographical coordinate system. The resolution used in this research is 30 meters with the extents of -7.748274 (*up*), 114.2937 (*left*), 114.468 (*right*), -7.928792 (*bottom*). Data extraction techniques environmental variables that influence habitat suitability *Acacia nilotica* can be seen in Table 1. Data processing is done by using Libre Office Calc, Quantum GIS, and R Studio.

2.3. Overlay Data

The overlay is done to combine several environment variables that are used into one predictor variable. The result of the overlay can be used to extract the values of the predictor variable. In addition, the overlay results into one of the components needed to build predictive models.

2.4. Testing Multicollinearity

Before doing the modeling, multicollinearity test needs to be done. Multicollinearity is a condition where there is a very strong relationship between predictor variables. Multicollinearity test is performed to determine whether there is collinearity between predictor variables. Mathematically multicollinearity can mainly be detected with the help of tolerance and its reciprocal, called variance inflation factor (VIF). Tolerance approaching 1 can be indicated that the value of multicollinearity is very small, in contrast, if tolerance close to 0 can indicate the occurrence of multicollinearity in that variable [21]. The calculation of the value of VIF to 6 environment variables according to (1).

$$VIF=\frac{1}{1-R^2} \tag{1}$$

Where R^2 is the coefficient determinant of the predictor variable.

2.5. Data Partition

The spread of data in the form of coordinate points is then divided into training data and testing data. The technique used in data dividing is using K-fold Cross Validation. K-fold Cross Validation is a method used to evaluate learning algorithm by dividing data into k-fold since k-1 fold is used as training data and 1 fold is used as data test [22]. Training data is used to build predictive models, while test data is used to test the performance of the model.

2.6. Maximum Entropy

The availability of data on environmental variables that affect species existence as well as the development of supporting technologies in the data processing has led to the development of predictive modeling based on environmental factors and the existence of species geographically [17]. For some species that have data and information on presence and absence, it is possible to use some statistical techniques in predictive modeling. However, not all species have the data and complete information on the presence and absence, so it requires a certain modeling method for predicting the presence of such species geographically [23].

Maxent is a general purpose method for making predictions from incomplete information. Maxent is a common approach for modeling species distribution using only the presence dataset. The idea of Maxent is to estimate a target probability distribution by finding the probability distribution of maximum entropy, subject to a set of constraints that represent our incomplete information about the target distribution [23]. Basically, the problem of modeling species distribution is a matter of density estimation. In the estimation, the maximum entropy density of the species distribution is represented by the probability distribution π above the set X of the study location. Thus, π gives a non-negative value for each x and the sum of values π(x) is 1 [24]. If we assume the response variable as y, then π(x) is the conditional probability $P(x \mid y{=}1)$, that is the probability of evidence x, if the hypothesis y=1 is known. According to Bayes' rule shown by (2) [24].

$$P(y{=}1 \mid x)=\frac{P(x \mid y{=}1)\,P(y{=}1)}{P(x)}=\pi(x)P(y{=}1)|X| \tag{2}$$

Where,

y =class target
x =evidence
P(x) =probability of evidence x
P(y=1) =probability of the number of data with the presence class
P(x | y=1) =probability of presence
P(y=1 | x)=probability of data with the presence class, if given the evidence x

Equation (2) showed that the π is comparable to the probability of presence. However, if it has only presence data, it cannot determine the probability of species presence [23, 25]. Therefore we estimate the distribution of "π" before making an estimate of $P(y{=}1 \mid x)$.

Presented a new way to estimate the probability distribution targets of the model Maxent. The probability distribution target can be calculated using the Gibbs Distribution theorem. According to [26], the Maxent distribution belongs to the family of Gibbs distributions derived from the set of features $f_1,\ldots,f_n$. Gibbs distribution is exponential distribution parameterized by a vector of feature weights $\lambda=(\lambda_1,\ldots,\lambda_n)$, and is defined (3) [24].

$$q_\lambda(x)=\frac{\exp\left(\sum_{j=1}^{n}\lambda_j f_j(x)\right)}{Z_\lambda} \tag{3}$$

Where Z_λ is a normalization constant ensuring that probabilities $q_\lambda(x)$ sum to one over the study area. After obtaining an estimate of q_λ, sufficient information is obtained to obtain the probability distribution P(y=1|x), as indicated (4) [24].

$$P(y{=}1|x)=\frac{e^H q_\lambda(x)}{1+e^H q_\lambda(x)} \tag{4}$$

Where q_λ is the estimated probability of presence with the maximum entropy of π and H is the entropy of q_λ.

2.7. Calibration and Evaluation

Some studies have independent data available for validation [27-28] or collect new data for validation [29]. However, it is often not feasible to collect new data. In this situation, a very common approach used to validate is to divide the data into a single section used for model calibration (training data), and one other section used for model validation (test data) [30, 31].

In this study, the distribution of data using K-fold cross validation. K-fold cross validation is a method to evaluate learning algorithm performance by dividing data into k fold, as much as k-1 fold is used as training data and 1 fold as test data [22].

In modeling the distribution of species, evaluating the habitat suitability model and the resulting predictive maps focused on quantifying predictive accuracy as a model performance measure or validity [32]. The species distribution model is validated using new data or independent data [33]. Using the same data to calibrate and evaluate the species distribution model would result in excessive model performance [34]. Most of the modeling methods used in the species distribution model to assign classifications, actually predict the probability of class membership. Conventionally, probabilistic predictions are converted to categorical by using a threshold probability value to distinguish presence and absence [32].

3. Results and Analysis

3.1. Model Performance and Variable Contribution

The result of the model evaluation in predicting the existence of *Acacia nilotica* in the Baluran National Park can be seen from one of the Maxent output, graphics of sensitivity and 1-specificity is shown in Figure 2.

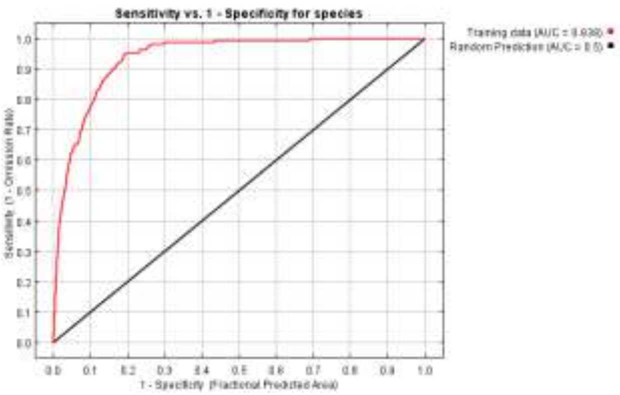

Figure 2. ROC curve

Those graphics shows the AUC score of the training data is 0.938. According to [35], models can be categorized as providing good performance. The result of variables contribution tested using jackknife is shown in Figure 3.

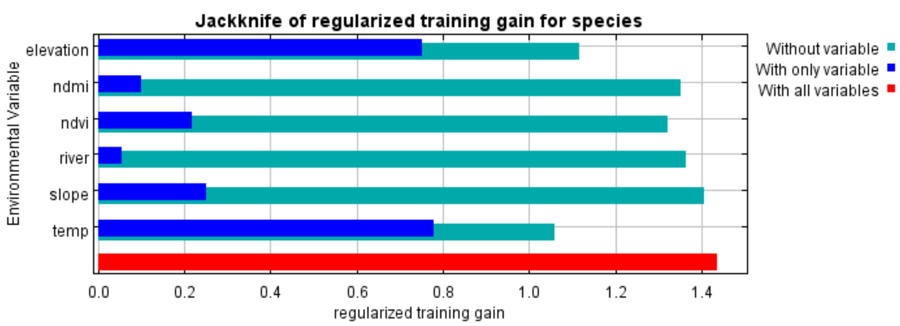

Figure 3. The Test results with the jackknife variable contribution in modeling the spread of *Acacia Nilotica*

The variables of temperature and elevation produce the high gain (> 0.6) if they are used independently. That information shows that temperature and elevation have information that is more useful than the other variables. The variable of elevation is an important fact which having effects on the other plants [36]. NDMI, NDVI, and Slope have a medium gain if it used independently. Therefore, the variable of distance from the river has a low gain if it used independently which shows that this variable does not have a lot of useful information.

3.2. Habitat Suitability Response Curves

The relationship between the probability of *A. nilotica* presence and environmental variable can be seen in response curves produced by Maxent. The curves show how the varied environmental variable effect on the prediction of *A. nilotica* presence as shown in response curves from 6 environmental variables in Figure 4. Response curves show the quantitative relationship between environmental variable with the logistic probability of presence (also known wit habitat suitability) and deepen the understanding of niche ecology of the species [16].

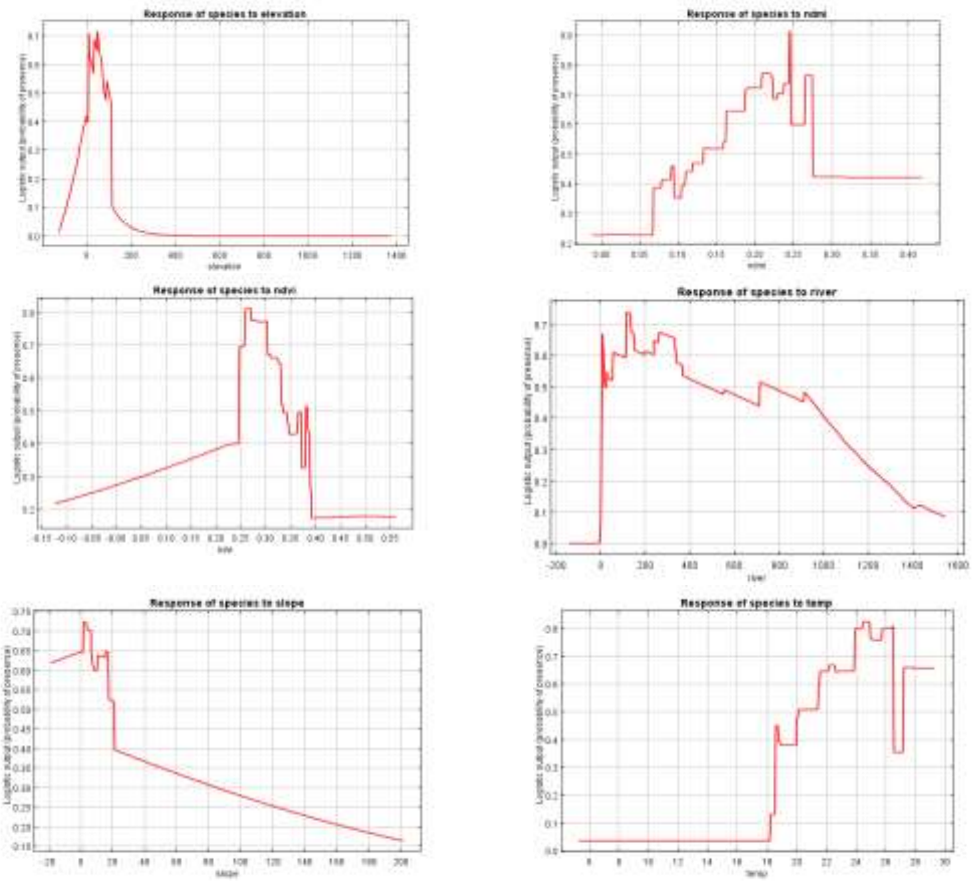

Figure 4. Response curves of the six variables in the model predictions of the spread of *Acacia Nilotica*

3.2.1. Habitat Suitability based on Elevation

According to the response curves, a suitable elevation range is 0-100 meters, included in the category of flat topography and the flat savanna with soil deposits (alluvial) [1]. *A. nilotica* thrives on alluvial soil with a high clay levels [1, 37] and can grow on poor soil nutrient elements [1].

3.2.2. Habitat Suitability based on NDMI

In the response curves in NDMI produce score NDMI for about 0-0.45. The score which closer to 1 describes that the area is flooded by water often or nearby the river. The curves describe the probability of *A. nilotica* presence in NDMI for about 0.15-0.25 and then decrease in the index of more than 0.25. These results indicate that A. nilotica grows fertile in dry areas [37].

3.2.3. Habitat Suitability based on NDVI

NDVI gained in the modeling which then shown by the response curves is about -0.15-0.55 according to the statement of Dragomir, Petrosani [38], that the NDVI score about -1-1. Next, the results of our research are suitable which the previous research which is done by Siswoyo [18], which declared that the result of calculation produces NDVI score between 0.1147-0.5243. The probability shown by the response curves in NDVI variable shows that the land covering/vegetation in the Baluran National Park is a Savanna, secondary forest mix with scrub and primary forest.

3.2.4. Habitat Suitability based on Distance from the River

The model predicts the decrease of suitability with the raising distance to the nearest river. It shows that the *A. nilotica* usually lives in a flat flood plain that suitable to the research Duke [37], which declared that *A. nilotica* generally grows near the waterway, especially in the flood area.

3.2.5. Habitat Suitability based on Slope

In the response curves of the slope, variable produce score 0-18 degrees which mean that the habitat of *A. nilotica* is in a flat and sloping topography [1]. It is closely related to the diverse behaviour of the herbivores which become the seeds disseminators [18]. Herbivorous animals favour low-tilt habitats aimed at meeting all life needs such as eating and drinking [9].

3.2.6. Habitat Suitability based on Temperature

The temperature response curves show the logistic probability of *A. nilotica* presence increases started from 18.5^{o}C-26.5^{o}C and then decrease. The result of the curves shows the condition that suitable with the research [39], which declared that the annual average temperature that suitable with the habitat of *A. nilotica* is between 18.7^{o}C-27.8^{o}C.

3.3. Model Application

3.3.1. The *Acacia nilotica* Distribution

The distribution of *Acacia nilotica* in Baluran National Park is shown in Figure 5. The temperature variable is the most important variable structuring *A. nilotica* distribution. As the temperature increases, the suitability of habitat increases. The elevation variable is the variable that influences after the temperature. Because the habitat of *A. nilotica* tends to be on the ground is not too high, in other words in the range of 0-100 meters. Furthermore, the vegetation index and moisture index are two variables that influence. Suitable habitats are distributed in open areas or have a vegetation index of less than 0.5 and are slightly waterlogged or have a humidity index between 0.15-0.25. The slope and distance from the nearest river become the next important variable.

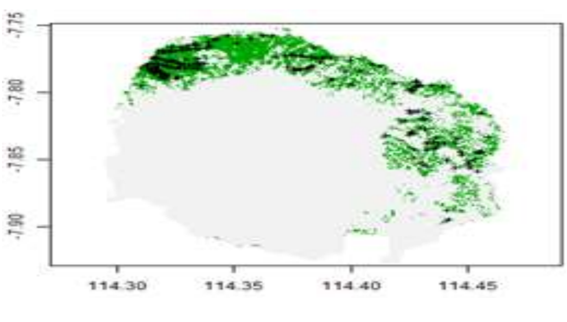

Figure 5. Habitat suitability distribution of *A. Nilotica* according to occurrence records

3.4. Calibration and Evaluation

The calibration and evaluation are done by dividing the presence data using the K-fold method. Data is divided into 10 fold. A total of 9 fold is used as training data and 1 fold is used as test data. The evaluation phase is crucial in assessing the accuracy of predictions. This is achieved by testing the potential distribution of a species represented by the prediction model against evidence recorded in the field [40]. In the evaluation phase, the additional data used is the absence data obtained from the observation [18]. The results of the evaluation show excellent model performance as shown in Figure 6. Figure 6 shows that model performance is categorized very well with an AUC value of 0.984. It shows that the test data used can represent the whole data in the study area.

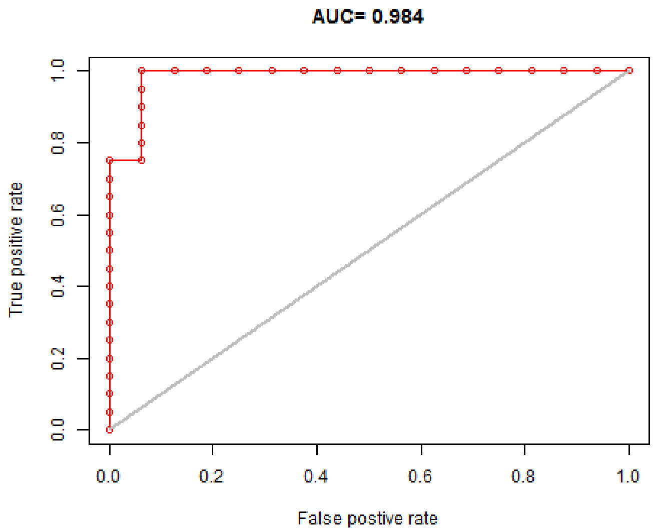

Figure 6. ROC curve evaluation

4. Conclusion

A model of the prediction of the potential presence based on the maximum entropy theory is developed to evaluate and predict the potential existence of *Acacia nilotica* in the Baluran National Park. Predictive models can be categorized very well as indicated by the AUC value of 0.938. Environmental factors that most influence the frequency of presence of *A. nilotica* is the elevation and temperature. It explains that *A. nilotica* has a habitat in the lowlands and medium temperature. Most of the research sites are considered to have low suitability as a habitat for *A. nilotica*. For further research, the model can be developed by adding the geological factor in the form of soil variable with the categorical data type.

References

[1] Sabarno MY. Savana Taman Nasional Baluran. *Biodiversitas*. 2002; 3(1): 207-12.
[2] Gunaryadi D. Pengamatan Populasi Cervus timorensis di Savana Bekol Taman Nasional Baluran Jawa Timur. Disertasi; 1996.
[3] Qirom MA, Andriani S, Azwar F, Octavia D. Pengaruh Pembebasan Jenis Akasia Berduri Acacia nilotica (L.) Willd. ex Del terhadap Komposisi Jenis Tumbuhan Penyusun Savana dan Kualitas Savana di Taman Nasional Baluran, Jawa. *Jurnal Penelitian Hutan dan Konservasi Alam*. 2016; 4(6): 573-82.
[4] Brenan JPM. Manual on Taxonomy of Acacia Species. Present Taxonomy of Four Species of Acacia (A. albida, A. senegal, A. nilotica, A. tortilis). *Manual on taxonomy of Acacia species Present taxonomy of four species of Acacia (A albida, A senegal, A nilotica, A tortilis)*. 1983.
[5] Gupta R. Resource Survey of Gummiferous Acacias in Western Rajasthan. *Tropical Ecology*. 1970; 11(2): 148-61.
[6] BTNB. *Rancangan Pencabutan Seedling/Anakan Hasil Pembongkaran secara Mekanis, 150 ha di Savana Bekol Taman Nasional Gunung Baluran*. Balai Taman Nasional Baluran. 1999.

[7] Djufri. Acacia nilotica (L.) Willd. ex Del. dan Permasalahannya di Taman Nasional Baluran Jawa Timur. *Biodiversitas*. 2004; 5: 96-104.
[8] Elith J, Leathwick JR. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual review of ecology, evolution, and systematics*. 2009; 40: 677-697.
[9] Rahmat UM, Santosa Y, Prasetyo LB, Kartono AP. Habitat Suitability Modeling of Javan Rhino (Rhinoceros sondaicus Desmarest 1822) Ujung Kulon National Park. *Jurnal Manajemen Hutan Tropika*. 2012; 18(2): 129-37.
[10] Prayogo H, Thohari AM, Solihin DD, Prasetyo LB, Sugardjito J. Habitat Suitability Modeling of Bornean Orangutan (Pongo pygmaeus pygmaeus) in Betung Kerihun National Park, Danau Sentarum and Corridor, West Kalimantan. *Jurnal Manajemen Hutan Tropika*. 2014; 20(2): 112-20.
[11] Favata CA, Christensen DR, Thompson R, McKeown KA, Hanselman JA. Evaluation of a Modified Habitat Suitability Index Model for Eastern Brook Trout: Implications for Efficient Habitat Assessment. *Journal of Student Research*. 2015; 4(1): 90-8.
[12] Prasetyo LB, Supartono T, Kartono AP, Hikmat A, Ramdhoni S. *Habitat Suitability Index (HIS) of Surili (Presbytis comata Desmarest, 1822) in Mixed Forest of Kuningan District, West Java-Indonesia*. IOP Conference Series: Earth and Environmental Science. 2017;54: 012061.
[13] Guisan A, Edwards TC, Hastie T. Generalized Linear and Generalized Additive Models in Studies of Species Distributions: Setting the Scene. *Ecological modelling*. 2002; 157(2): 89-100.
[14] Ulrich W, Soliveres S, Thomas AD, Dougill AJ, Maestre FT. Environmental Correlates of Species Rank−abundance Distributions in Global Drylands. *Perspectives in plant ecology, evolution and systematics*. 2016; 20: 56-64.
[15] Bowler M. Species Abundance Distributions, Statistical Mechanics and the Priors of MaxEnt. *Theoretical population biology*. 2014; 92: 69-77.
[16] Yi YJ, Cheng X, Yang ZF, Zhang SH. Maxent Modeling for Predicting the Potential Distribution of Endangered Medicinal Plant (H. riparia Lour) in Yunnan, China. *Ecological Engineering*. 2016; 92: 260-9.
[17] Rusman D. Prediksi Kehadiran Badak Sumatera (Dicerorhinus sumatrensis) dan Analisis Struktur Lanskap Habitatnya di Taman Nasional Bukit Barisan Selatan. Universitas Gadjah Mada Yogyakarta 2016.
[18] Siswoyo A. Pemodelan Spasial Kesesuaian Habitat Akasia Berduri (Acacia nilotica) di Taman Nasional Baluran. Thesis. Bogor: Bogor Agricultural University; 2014.
[19] Zhigang Z, Jingqin W, Li W, Meng W. Reliability Evaluation of Low-voltage Switchgear Based on Maximum Entropy Principle. *TELKOMNIKA (Telecommunication, Computing, Electronics and Control)*. 2017; 15(1).
[20] Kumar YJN, Kanth TR. GIS-MAP Based Spatial Analysis of Rainfall Data of Andhra Pradesh and Telangana States Using R. *International Journal of Electrical and Computer Engineering (IJECE)*. 2017; 7(1): 460.
[21] Midi H, Sarkar S, Rana S. Collinearity Diagnostics of Binary Logistic Regression Model. *Journal of Interdisciplinary Mathematics*. 2010; 13(3): 253-67.
[22] Liu L, Özsu MT. Encyclopedia of Database Systems: Springer Berlin, Heidelberg, Germany. 2009.
[23] Phillips SJ, Anderson RP, Schapire RE. Maximum Entropy Modeling of Species Geographic Distributions. *Ecological modelling*. 2006; 190(3): 231-59.
[24] Phillips SJ, Dudík M. Modeling of Species Distributions with Maxent: New Extensions and a Comprehensive Evaluation. *Ecography*. 2008; 31(2): 161-75.
[25] Ward G, Hastie T, Barry S, Elith J, Leathwick JR. Presence-only Data and the EM Algorithm. *Biometrics*. 2009; 65(2): 554-63.
[26] Dudik M, Phillips SJ, Schapire RE. *Performance Guarantees for Regularized Maximum Entropy Density Estimation*. International Conference on Computational Learning Theory. 2004; 472-86.
[27] Elith* J, H. Graham* C, P. Anderson R, Dudík M, Ferrier S, Guisan A, et al. Novel Methods Improve Prediction of Species' Distributions from Occurrence Data. *Ecography*. 2006; 29(2): 129-51.
[28] Franklin J. Enhancing a Regional Vegetation Map with Predictive Models of Dominant Plant Species in Chaparral. *Applied Vegetation Science*. 2002; 5(1): 135-46.
[29] Fielding AH, Bell JF. A Review of Methods for the Assessment of Prediction Errors in Conservation Presence/Absence Models. *Environmental conservation*. 1997; 24(1): 38-49.
[30] Smith P. Autocorrelation in Logistic Regression Modelling of Species' Distributions. *Global ecology and biogeography letters*. 1994: 47-61.
[31] Miller J, Franklin J. Predictive Vegetation Modeling with Spatial Dependence Vegetation Alliances in the Mojave Desert. *Ecological Modelling*. 2002; 157: 225-45.
[32] Franklin J. Mapping Species Distributions: Spatial Inference and Prediction: Cambridge University Press. 2010.
[33] Barry S, Elith J. Error and Uncertainty in Habitat Models. *Journal of Applied Ecology*. 2006; 43(3): 413-423.

[34] Edwards TC, Cutler DR, Zimmermann NE, Geiser L, Moisen GG. Effects of Sample Survey Design on the Accuracy of Classification Tree Models in Species Distribution Models. *ecological modelling*. 2006; 199(2): 132-141.
[35] Swets JA. Measuring the Accuracy of Diagnostic Systems. *Science*. 1988; 240(4857): 1285-93.
[36] Adhikari D, Barik S, Upadhaya K. Habitat Distribution Modelling for Reintroduction of Ilex khasiana Purk., a Critically Endangered Tree Species of Northeastern India. *Ecological Engineering*. 2012; 40: 37-43.
[37] Duke J. Handbook of Legumes of World Economic Importance. New York: Plenum Press. 2012.
[38] Dragomir LO, Petrosani PDECH, Oncia S. Using Satellite Images Landsat TM for Calculating Normalized Difference Indexes for The Landscape of Parang Mountains. GeoCAD; 2012.
[39] Ruskin F. Firewood Crops. Shrub and Tree Species for Energy Production. *Firewood crops Shrub and tree species for energy production*. 1980.
[40] Ottaviani D, Lasinio GJ, Boitani L. Two Statistical Methods to Validate Habitat Suitability Models using Presence-only Data. *Ecological Modelling*. 2004; 179(4): 417-43.