

# Pemodelan *Ensemble* Prediksi Distribusi Ekologis Padi (*Oryza sativa*) di Provinsi Kalimantan Utara.

Patria Kusumadiya<sup>1\*</sup>, Omo Rusdiana<sup>2</sup>, dan Sri Mulatsih<sup>3</sup>

<sup>1</sup>Program Studi Ilmu Pengelolaan Sumberdaya Alam dan Lingkungan, Sekolah Pascasarjana, IPB University, Bogor 16144, Indonesia;

<sup>2</sup>Departemen Silvikultur, Fakultas Kehutanan dan Lingkungan, IPB University, Bogor 16680, Indonesia

<sup>3</sup>Departemen Ilmu Ekonomi, Fakultas Ekonomi dan Manajemen, IPB University, Bogor 16680, Indonesia

## ABSTRAK

Pemerintah Provinsi Kalimantan Utara berusaha mencapai ketahanan pangan dengan prinsip kemandirian pangan melalui perluasan lahan pertanian. Penilaian kesesuaian lahan pertanian, terutama untuk padi, dilakukan menggunakan pendekatan pemodelan ensemble yang melibatkan lima algoritma pembelajaran mesin. Model-model ini dibangun menggunakan paket *species distribution modeling* (SDM) di RStudio dengan pembagian data pelatihan dan pengujian 70:30 serta pengaturan parameter termasuk bootstrapping dan tiga kali pengulangan. Hasil penelitian menunjukkan variasi dalam respons variabel prediktor antara algoritma. Variabel NDVI memiliki pengaruh tertinggi pada SVM dan BRT (masing-masing 48,1% dan 36,6%), sementara variabel jarak dari jalan paling berpengaruh pada GLM, MARS, dan RF (masing-masing 44,6%, 27,6%, dan 26,5%). Distribusi padi (sawah) bervariasi antara model, dengan RF memiliki persentase tertinggi (6,34%). Evaluasi kinerja model-model ini menunjukkan bahwa model RF memiliki akurasi terbaik, sementara GLM memiliki akurasi buruk dalam nilai Kappa Cohen. Model ensemble memperoleh akurasi yang dapat diterima dengan nilai masing-masing 0,96; 0,70; dan 0,71 untuk AUC, TSS, dan Kappa. Dengan demikian, pendekatan pemodelan multi-algoritma dengan model ensemble memungkinkan penilaian yang lebih baik terhadap variabilitas dalam kinerja algoritma dan menghasilkan peta kesesuaian distribusi padi yang lebih baik daripada algoritma tunggal.

**Kata kunci:** *Ensemble*, Model distribusi spesies, Regresi, Pembelajaran mesin, Kesesuaian lahan padi.

## ABSTRACT

The North Kalimantan Provincial Government is striving to achieve food security with the principle of food self-sufficiency through the expansion of agricultural land. The assessment of agricultural land suitability, especially for rice cultivation, is carried out using an ensemble modeling approach that involves five machine learning algorithms. These models are built using the *species distribution modeling* (SDM) package in RStudio with a 70:30 split of training and testing data and parameter settings that include bootstrapping and three repetitions. The research results show variation in the predictor variable response among algorithms. The NDVI variable has the highest influence on SVM and BRT (48.1% and 36.6%, respectively), while the distance from the road has the most significant impact on GLM, MARS, and RF (44.6%, 27.6%, and 26.5%, respectively). The distribution of rice fields varies among models, with RF having the highest percentage (6.34%). The performance evaluation of these models indicates that the RF model has the highest accuracy, while GLM performs poorly in terms of Kappa Cohen's value. The ensemble model achieves acceptable accuracy with respective values of 0.96, 0.70, and 0.71 for AUC, TSS, and Kappa. Thus, a multi-algorithm modeling approach with ensemble models allows for a better assessment of algorithm performance variability and produces a more accurate suitability map for rice distribution than a single algorithm.

**Keywords:** Ensemble, Species distribution model, Regression, Machine learning, Rice field suitability

**Citation:** Kusumadiya, P., Rusdiana, O., dan Mulatsih, S., (2024). Pemodelan Ensemble Prediksi Distribusi Ekologis Padi (*Oryza sativa*) di Provinsi Kalimantan Utara. *Jurnal Ilmu Lingkungan*, 22(2), 313-325, doi:10.14710/jil.22.2.313-325

## 1. Pendahuluan

Secara global, beras ditanam di berbagai wilayah di dunia: Asia Tenggara, Asia Selatan, negara-negara Mediterania, Timur Tengah, Afrika, Amerika Serikat, Brasil dan beberapa negara bagian Amerika Latin lainnya (Yoshida, 1981; GRiSP, 2013). Telah terungkap bahwa beras adalah tanaman paling

penting yang memberi makan lebih banyak orang daripada tanaman lain di dunia (GRiSP, 2013; Muthayya et al., 2014).

Berdasarkan data BPS Kalimantan Utara (2020), tahun 2019 mencatatkan produksi padi sebanyak 33.357,19 ton atau setara dengan 17.919,72 ton beras, jumlah ini lebih rendah dibandingkan dengan tahun

2018 yang mencapai 45.063,53 ton padi atau setara dengan 24.208,45 ton beras. Penurunan luas panen tersebut lebih dipengaruhi oleh penurunan produktivitas daripada penurunan luas panen. Meskipun begitu, kebutuhan pangan berupa beras untuk penduduk Kalimantan Utara yang berjumlah 742.200 jiwa pada tahun 2019 masih jauh dari terpenuhi. Hal ini terlihat dari konsumsi beras per kapita yang mencapai 111,58 kg per tahun (BPS, 2019). Dengan demikian, pada tahun 2019, produksi beras Kalimantan Utara hanya cukup untuk memenuhi kebutuhan sekitar 193.651 jiwa.

Sebagai upaya mewujudkan ketahanan pangan, pemerintah provinsi merumuskan 10 program prioritas sebagai langkah percepatan pembangunan, salah satu program prioritasnya adalah ketahanan pangan dengan prinsip kemandirian pangan. Aziza et al. (2008) menyebutkan bahwa perwujudan ketahanan pangan pada subsistem produksi salah satunya dapat ditempuh melalui perluasan lahan (pertanian) atau ekstensifikasi pertanian.

Rencana perluasan dan pemanfaatan atau pengelolaan lahan harus tetap memperhatikan kesesuaian atau kemampuan lahan (KLH, 2009). Oleh karena itu, penting untuk memproyeksikan kesesuaian lahan untuk penanaman padi sebagai upaya mengembangkan strategi adaptasi dan mitigasi yang tepat untuk ketahanan pangan. Secara kritis, menjaga dan merehabilitasi habitat sawah juga berkontribusi pada perlindungan dan konservasi ekosistem (Dang et al., 2020)

Baru-baru ini, berbagai program pemodelan telah dikembangkan untuk menganalisis kesesuaian lahan untuk spesies alami dan komunitas ekologi dan produksi tanaman atas pengaruh perubahan iklim berdasarkan pendekatan *GIS* dan penginderaan jauh. Salah satu diantaranya, *species distribution modeling* (SDM) sebagai alat empiris memiliki kapasitas untuk memprediksi distribusi spesies dengan menghubungkan data spesies geo-lokal dengan variabel lingkungan (Byeon, 2018)

Penggunaan *species distribution modeling* (SDM) dapat membantu tidak hanya dalam mengidentifikasi kondisi yang cocok untuk keberadaan spesies, tetapi juga dalam memprediksi distribusi spasial (Fourcade et al., 2014; Remya et al., 2015). Sehingga, SDM telah banyak digunakan dalam manajemen sumber daya alam dan ekologi (Byeon, 2018), seperti penilaian risiko spesies invasif, prediksi keanekaragaman hayati dan pembentukan lokasi yang cocok untuk budidaya spesies di bawah berbagai skenario perubahan iklim (Austin, 2002; Hossell et al., 2003; Lamsal et al., 2018; Kariyawasam, 2019). Akhir-akhir ini, banyak penelitian telah menunjukkan potensi *species distribution modeling* dalam memprediksi potensi distribusi tanaman pertanian (Davis et al., 2012; Duan dan Zhou, 2012; Liu et al., 2015; He dan Zhou, 2016; Jalaiean et al., 2018) atau dampak perubahan iklim pada tanaman pertanian dengan menggabungkan SDM dengan model sirkulasi umum (GCM) [Davis et al., 2012; Kogo et al., 2019; Layomi

Jayasinghe et al., 2019; Ratnayake et al., 2019; Chhogyel et al., 2020).

Penerapan model algoritma tunggal dalam literatur SDM telah menimbulkan perdebatan, karena variasi kinerja prediktif yang sangat tinggi jika dibandingkan dengan model-model lainnya (Thuiller et al., 2019). Früh et al. (2018) melakukan perbandingan kinerja empat model pembelajaran mesin dalam memprediksi spesies invasif, dan mereka merekomendasikan pendekatan ensemble dari model dengan kinerja terbaik daripada mengandalkan model tunggal atau *ensemble* dari semua model. Hasil perbandingan antara model pembelajaran mesin menunjukkan bahwa model *Random Forest* (RF) dan *ensemble* sangat dihargai dibandingkan dengan model lainnya (Ng et al., 2018). RF merupakan algoritma pembelajaran mesin yang menggunakan kombinasi pohon prediktor dan terbukti efektif dalam memprediksi kekayaan dan kepadatan spesies secara menyeluruh (Kosicki 2020).

Abdi (2020) juga melakukan perbandingan empat algoritma pembelajaran mesin dengan menggunakan variabel turunan S2 untuk klasifikasi tutupan lahan, dan ia menyimpulkan bahwa *Support Vector Machine* (SVM) memiliki kinerja lebih baik daripada model-model lainnya. Algoritma SVM mengubah ruang asli dan kemudian membangun hiperplane optimal dalam ruang fitur multi-dimensi, yang memisahkan data ke dalam kelas-kelas yang berbeda. BRT yang bergantung pada jumlah pohon yang relatif kecil untuk meningkatkan kinerja variabel prediktif dan memiliki kemampuan untuk memproses beberapa prediktor dengan akurasi tinggi, juga dianggap sebagai salah satu pilihan yang baik (Gu et al., 2019). Sementara itu, GAM adalah model regresi yang populer dan sering digunakan dalam studi ekologi untuk memodelkan distribusi yang tidak normal (Ravindra et al., 2019). MARS adalah suatu model yang memiliki kemampuan untuk mengasumsikan bentuk hubungan fungsional antara variabel yang direspons dan variabel prediktor yang tidak diketahui, serta model ini memiliki fleksibilitas dalam hal bentuk fungsionalnya. Penelitian-penelitian ini melibatkan penggabungan data dari beragam sumber, termasuk data dari penginderaan jauh dan data non-penginderaan jauh (Ng et al., 2018).

Manfaat *Random Forest* (RF) dapat dilihat dari kemampuannya untuk mengatasi overfitting dan mengatasi sejumlah besar variabel prediktor (Ng et al., 2018). Di sisi lain, model *Generalized Linear Model* (GLM) cocok untuk memodelkan hubungan linier antara variabel. *Support Vector Machine* (SVM) memiliki keunggulan dalam menangani data yang bising dan jumlah sampel pelatihan yang terbatas dengan tingkat akurasi yang tinggi. *Generalized Additive Model* (GAM) memiliki fleksibilitas dalam menangani hubungan kompleks dan non-linier antar variabel, yang seringkali sulit diakomodasi oleh model regresi linier tradisional. Sementara itu, *Multivariate Adaptive Regression Splines* (MARS) dengan mudah menangani variabel kategori dan tidak memerlukan

konversi menjadi variabel numerik (Tomala et al., 2020).

Namun, terdapat beberapa keterbatasan dalam penggunaan masing-masing metode. RF dapat menjadi kompleks dan sulit diinterpretasikan, sedangkan GLM mungkin tidak cocok untuk model hubungan yang non-linier. SVM tidak toleran terhadap data yang bising dan dapat menjadi sensitif terhadap skala variabel. GAM dan MARS juga dapat menghadapi potensi overfitting jika tidak dikelola dengan baik (Raczko dan Zagajewski, 2018). Kombinasi RF, GLM, MARS, CVM, dan BRT dapat digunakan untuk mencakup berbagai jenis hubungan data atau mengatasi kekurangan satu metode dengan yang lain (Ahmed et al., 2021).

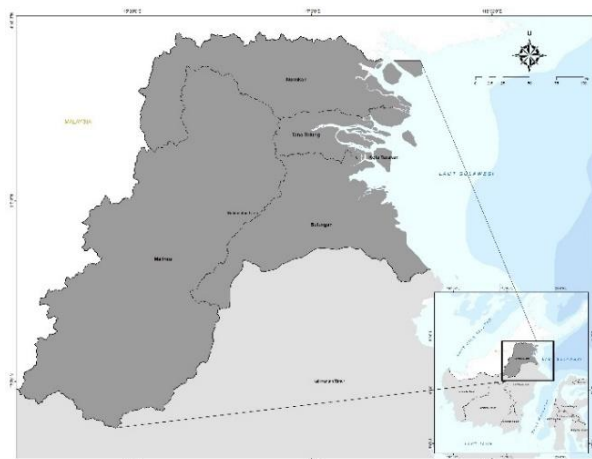
Oleh karena itu, disarankan untuk menggunakan pendekatan multi model, khususnya model *ensemble*, untuk meningkatkan *robustness* prediksi model dan interpretasinya (Araújo dan New, 2007). Studi terbaru, telah membandingkan pendekatan model *ensemble* dengan model algoritma tunggal, telah menunjukkan bahwa prediksi model *ensemble* secara signifikan berkinerja lebih baik dibandingkan dengan model tunggal (Mugo dan Saitoh, 2020). Penggunaan SDM, khususnya pendekatan pemodelan *ensemble* belum digunakan terutama sebagai prediksi kesesuaian lahan di Provinsi Kalimantan Utara.

Oleh karena itu, berdasarkan latar belakang tersebut penelitian ini bertujuan sebagai berikut: 1) mengetahui variabel penting dari variabel lingkungan yang digunakan sebagai input model terhadap *niche* (relung ekologi) vegetasi padi dibudidayakan di Kalimantan Utara; 2) menganalisis sebaran lokasi pengembangan/perluasan lahan pertanian di Kalimantan Utara dari hasil menilai potensi sebaran/distribusi vegetasi (padi) yang berasal dari beberapa variabel biofisik dan infrastruktur kunci untuk pemodelan distribusi spesies. Temuan penelitian ini akan memberikan informasi penting untuk mengembangkan strategi manajemen dan adaptasi pengembangan lahan pertanian berdasarkan parameter lingkungan yaitu digunakan sebagai prediksi kesesuaian lahan padi di Kalimantan Utara

## 2. Metode

### 2.1. Lokasi Penelitian

Penelitian ini dilakukan pada seluruh kabupaten, wilayah administrasi Provinsi Kalimantan Utara (Gambar 1), secara geografis terletak antara 114° 35' 22"- 118° 03' 00" BT dan antara 1° 21' 36"- 4° 24' 55" LU. Provinsi Kalimantan Utara berbatasan langsung dengan Sabah, Malaysia pada sisi Utara, dengan Provinsi Kalimantan Timur pada sisi Selatan, dengan Sarawak, Malaysia pada sisi Barat, dan sisi timur dengan Sulawesi. Provinsi Kalimantan Utara terbagi menjadi 5 (lima) wilayah administrasi kabupaten/kota yaitu, Kabupaten Tana Tidung, Kabupaten Bulungan, Kabupaten Nunukan, Kabupaten Malinau dan Kota Tarakan.



Gambar 1. Lokasi Penelitian

### 2.2. Pengumpulan Data

Jenis data yang digunakan, ukuran, sumber dan penggunaannya, ditunjukkan pada Tabel 1.

Tabel 1. Data yang digunakan, jenis dan penggunaannya

No	Data	Jenis/ Ukuran	Sumber	Penggunaan
1.	Landsat 8	Raster/ 30 m	GEE	Analisis Indeks Vegetasi (NDVI)
2.	Peta Tanah	Vektor/ Skala 1:50.000	BBSDLP Pertanian	Klasifikasi jenis tanah
3.	DEM	Raster/ 30 m	SRTM. NASA	Analisis dan klasifikasi kelas lereng dan ketinggian
4.	SKL Drainase Tahun 2014	Vektor/ Skala 1:50.000	Bappeda Kaltara	Klasifikasi SKL drainase
5.	SKL Bencana Tahun 2014	Vektor/ Skala 1:50.000	Bappeda Kaltara	Klasifikasi SKL bencana
6.	Jalan	Vektor/ Skala 1:50.000	INA Geoportal	Analisis Euclidian Distance
7.	Sawah	Vektor/ Skala 1:50.000	Kementerian Pertanian	Analisis Presence

### 2.3. Analisis Data

Data kehadiran vegetasi (padi) diperoleh dari melakukan *digitasi on screen* dengan panduan dari data vektor sawah tahun 2020 dari Kementerian Pertanian dan basemap citra resolusi tinggi Google Earth/Esri yang tersedia pada perangkat Arcgis. Jarak antar titik kehadiran vegetasi (padi) minimal adalah 100 m. Sedangkan, data ketidakhadiran diperoleh dari membuat *random point* menggunakan *Toolbox-Arcgis "Create Random Point"*, dengan jarak yang sama antar titik minimal 100 m.

Untuk mendapatkan performa model dilakukan evaluasi autokorelasi spasial di antara pengamatan, menggunakan indeks Moran's. Hasil Indeks Moran's menunjukkan tidak ada pengelompokan spasial yang jelas di antara poin-poin (Abdulhafedh, 2017). Selanjutnya, untuk lebih mengurangi autokorelasi spasial di antara titik-titik, menggunakan fungsi

"Rarefy Occurrence Data for SDMs (mengurangi autokorelasi spasial)" di ArcGIS SDM Toolbox. Untuk mendapatkan statistik validasi independen, 70% dari data yang dikumpulkan digunakan untuk melatih model sementara 30% digunakan untuk memvalidasi model (Engler et al., 2013).

Variabel lingkungan yang digunakan terbagi menjadi 2 (dua) kategori, pertama merupakan data kategorik (bifisik) terdiri dari 6 (enam) variabel data yaitu kelas lereng, kelas ketinggian, indeks vegetasi-NDVI (*Normalized Difference Vegetation Index*), satuan kemampuan lahan (SKL) bencana, SKL drainase, dan tanah. Kategori kedua merupakan data yang mewakili variabel infrastruktur berupa data kontinu yaitu jarak dari jalan.

Variabel kelas lereng merupakan hasil dari analisis *slope*, dan variable kelas ketinggian merupakan hasil dari analisis *reclassify* - toolbox ArcGIS. Variabel indeks vegetasi (NDVI) didapatkan dari proses analisis dari data Landsat 8. Untuk mendapatkan hasil indeks vegetasi (NDVI), menggunakan rumus:

$$NDVI = \frac{NIR-RED}{NIR+RED} \dots\dots\dots(1)$$

Keterangan :

NIR = Band 5

RED = Band 4

Proses analisis indeks vegetasi (NDVI) menggunakan *raster calculator-toolbox* ArcGIS. Variabel jarak dari jalan diperoleh dari analisis *Euclidean distance-toolbox* ArcGIS dengan input data dari data vektor jalan Rupa Bumi Indonesia (RBI). Sedangkan variabel SKL bencana, SKL drainase dan tanah, hanya dilakukan proses klasifikasi saja dikarenakan menggunakan data sekunder dari instansi pemerintah. Klasifikasi variabel lingkungan ditunjukkan pada Tabel 2.

Pra-pemrosesan seluruh data raster variabel lingkungan dilakukan untuk memperoleh data yang seragam meliputi kesamaan resolusi (*Cell Size*), dimensi (*nColoum* dan *nRow*), *extent* (Xmin, Xmax, Ymin, Ymax), dan *reference* (*Coordinate System*) data. Seluruh data variabel diproduksi dengan menggunakan perangkat lunak ArcGIS dengan *Extract by Mask* dan *Clip Raster* - toolbox ArcGIS.

Dalam memilih variabel yang akan digunakan dalam pembuatan model, dilakukan pengujian multikolinearitas. Pengujian ini bertujuan untuk menilai apakah ada hubungan kuat antara variabel-variabel tersebut. Pendekatan yang digunakan yaitu korelasi *variance inflation factor* (VIF) untuk mengurangi masalah multikolinearitas dalam paket *species distribution modeling "sdm"* perangkat lunak R 4.0 (Naimi dan Araújo, 2016). VIF digunakan oleh beberapa studi di bidang distribusi spesies sebagai alat untuk memilih variabel penting (Ng et al., 2018; Zheng et al., 2015). Selain itu, untuk mengetahui secara relatif variabel penting untuk semua model dihitung menggunakan fungsi "getVarImp" dalam

perangkat lunak R "SDM" paket (Naimi dan Araújo, 2016). Saat ini sejumlah besar metode pemodelan distribusi spesies yang tersedia, dapat diklasifikasikan dalam kelas model "profil (*profile*)," "regresi (*regression*)," dan "pembelajaran mesin (*machine learning*)" (Hijmans dan Elith, 2019).

**Tabel 2.** Klasifikasi data variabel lingkungan

Variabel	Minimum	Maksimum	Klasifikasi	Sumber
<i>Data kategorik (Biofisik)</i>				
Kelerengan	15 %	>15 %	1	Subroto dan Susetyo, 2016
	5 %	15 %	2	
	2 %	5 %	3	
	0 %	2 %	4	
Ketinggian	1000 mdpl	>1000 mdpl	1	Subroto dan Susetyo, 2016
	500 mdpl	1000 mdpl	2	
	100 mdpl	500 mdpl	3	
	0	100 mdpl	4	
NDVI	-1	≤ 0	1	Aquino et al., 2018
	0	≤ 0,2	2	
	0,2	≤ 0,4	3	
	0,4	≤ 0,6	4	
	0,6	≤ 0,8	5	
	0,8	≤ 1	6	
SKL Bencana	Tinggi	-	1	Kemen PU, 2007
	Sedang	-	2	
	Cukup	-	3	
	Kurang	-	4	
SKL Drainase	Kurang	-	1	Kemen PU, 2007
	Cukup	-	2	
	Sedang	-	3	
	Tinggi	-	4	
Tanah	Aluvial	-	1	Hasil Analisis, 2022
	Podsol	-	2	
	Podsolik	-	3	
	Regosol	-	4	
	Organosol	-	5	
	Oksisol	-	6	
	Mediterran	-	7	
	Litosol	-	8	
	Latosol	-	9	
	Gawir	-	10	
	Glei	-	11	
	Pulau-Pulau Kecil	-	12	
<i>Data kontinu (Infrastruktur)</i>				
Jarak ke Jalan	0 km	58,1409 km	-	Hasil Analisis, 2022

Dalam kajian ini, mengevaluasi kinerja 5 (lima) model yang umum digunakan di bidang SDM. Models dipilih dari algoritma pembelajaran mesin dan model regresi sebagai perbandingan (Tabel 3). *Boosted Regression Trees* (BRT), *Support Vector Machine* (SVM), dan *Random Forest* (RF), merupakan algoritma dari model pembelajaran mesin, sementara *Generalized Additive Model* (GAM), *Generalized Linear Model* (GLM), dan *Multivariate Adaptive Regression Spline* (MARS) adalah algoritma dari model regresi.

Ahmed et al., (2020) menyebutkan model distribusi spesies multi model (*Ensemble Models*), dengan mengkombinasikan pembelajaran mesin dan model regresi memiliki kapasitas yang lebih tinggi untuk memodelkan distribusi spesies.

**Tabel 3.** Algoritma model terpilih, deskripsi singkat dan literatur umum dalam pemodelan distribusi spesies.

Model	Deskripsi singkat	Referensi Contoh
<i>Random Forest</i> (RF)	RF adalah algoritma pembelajaran mesin (Abdi, 2020). RF adalah metode yang efektif dalam memprediksi kekayaan dan kepadatan spesies (Kosicki, 2020).	Früh et al., (2018); Mi et al., (2017); Ng et al., (2018); West et al., (2017)
<i>Support Vektor Machine</i> (SVM)	SVM (Cortes dan Vapnik 1995) mempunyai kemampuan untuk menggeneralisasi fitur.	Abdi (2020); Früh et al. (2018); Ng et al. (2016)
<i>Boosted Regression Trees</i> (BRT)	Seperti RF, BRT didasarkan pada kombinasi dari sejumlah kecil <i>Trees</i> untuk meningkatkan kinerja variabel prediktif (Elith et al., 2008). BRT juga memiliki kapasitas untuk memproses beberapa prediktor pada akurasi prediktif yang tinggi (Gu et al., 2019).	Guisan et al., (2007); West et al., (2017)
<i>General Linear Model</i> (GLM)	GLM dapat memproses dan mengelola struktur data non-linear. Fleksibilitasnya membuatnya lebih cocok untuk studi berbasis ekologis (Guisan et al., 2002).	Guisan et al., (2007); Soultan dan Safi, (2017); West et al., (2017)
<i>Multivariate adaptive regression spline</i> (MARS)	MARS merupakan satu model regresi nonparametrik, yaitu model yang dapat mengasumsikan bentuk hubungan fungsional antara variabel respon dan prediktor tidak diketahui, dan mempunyai bentuk fungsional yang fleksibel. (Friedman 1991).	Leathwick et al. (2006); Elith dan Leathwick (2007)

Selanjutnya, model *ensemble* dibangun menggunakan metode rata-rata tertimbang (*weighted mean*)-*True Skill Statistic* (TSS) dari semua model.

Untuk validasi model yang tercantum di atas, kami menggunakan pendekatan replikasi subsampling dan bootstrap dalam paket "SDM" perangkat lunak R yang dikembangkan oleh Naimi dan Araújo (2016). Dari total titik *presence-absence* yang dikumpulkan, (30%) titik yang dipilih secara acak digunakan sebagai data pengujian untuk memvalidasi model dan 70% sisanya digunakan untuk pelatihan model. Langkah ini direplikasi 3 (tiga) kali dan menghasilkan nilai rata-rata sensitivitas, spesifisitas, TSS, kappa, AUC, dan korelasi digunakan untuk menilai keakuratan model.

Metode replikasi bootstrapping memiliki potensi untuk menawarkan akurasi prediksi yang tidak bias dengan varians yang cukup rendah (Harrell et al., 1996; Lima et al., 2019). Selain itu, pendekatan maksimalisasi jumlah spesifisitas sensitivitas digunakan untuk memilih ambang batas terbaik. Ambang batas ini direkomendasikan sebagai pendekatan terbaik untuk prediksi distribusi spesies (Liu et al. 2005).

Peta biner dikembangkan sebagai piksel yang lebih besar dari ambang batas (*threshold*) yang mewakili keberadaan atau kesesuaian untuk spesies (padi) sawah, berbeda dengan piksel yang lebih rendah dari ambang batas menunjukkan tidak adanya atau tidak sesuainya untuk spesies padi (sawah) di daerah

tersebut untuk semua model.

Selain itu, model *ensemble* dievaluasi menggunakan mean tertimbang dari semua model dan Hasil model *ensemble* lebih lanjut diklasifikasikan menjadi "tidak diliputi," "rendah," "sedang," dan "tinggi" dari distribusi kesesuaian spesies padi (sawah) di Kalimantan Utara.

### 3. Hasil dan Pembahasan

#### 3.1. Pra-Pemrosesan Data

Total titik kehadiran dan ketidakhadiran (*presence-absence*) spesies (padi) sawah yang telah dilakukan analisis autokorelasi spasial di antara titik-titik, menggunakan fungsi "*Rarify Occurrence Data for SDMs*" adalah sejumlah 5039 titik. Terdiri dari 2542 titik kehadiran (*presence*) dan 2497 titik ketidakhadiran (*absence*). Sejumlah 30% titik (1512) digunakan sebagai data pengujian dan validasi model. Sedangkan 70% titik (3527) digunakan untuk pelatihan model.

Variabel lingkungan yang terdiri dari 7 (tujuh) data raster (Tabel 2) yang telah dilakukan proses klasifikasi, telah memiliki kesamaan resolusi (30 m), dimensi (n kolom: 9000 dan n baris: 8512), *extent* (Xmin: 339645; Xmax: 60945; Ymin: 232085,5; Ymax: 487445,5), dan koordinat sistem (UTM 48 S), ditunjukkan pada Gambar 2.

Dari hasil analisis VIF ditunjukkan pada Gambar 2, dari 7 (tujuh) variabel lingkungan (Tabel 2), total 6 (enam) variabel (selain SKL drainase) dengan nilai ambang batas kurang dari 0,7 terpilih (Engler et al., 2013). Dapat diartikan bahwa 6 (enam) variabel (selain SKL drainase) tidak terjadi permasalahan multikolinieritas antar variabel dan variabel tersebut akan digunakan dalam membangun model distribusi spesies.

#### 3.2. Kontribusi Relatif dari Variabel Prediktor

Pengaruh relatif variabel prediktor ditunjukkan pada Gambar 4. Pengaruh relatif dari variabel NDVI memiliki nilai tertinggi untuk SVM dan BRT masing-masing adalah 48,1% dan 36,6%. Pengaruh relatif dari variabel jarak dari jalan memiliki nilai tertinggi untuk GLM, MARS dan RF masing-masing adalah 44,6%, 27,6% dan 26,5%. Sedangkan rata-rata pengaruh relatif variabel dari kelima model, memiliki nilai tertinggi pada urutan pertama yaitu jarak dari jalan (28,3%) dan kedua NDVI (27,4%).

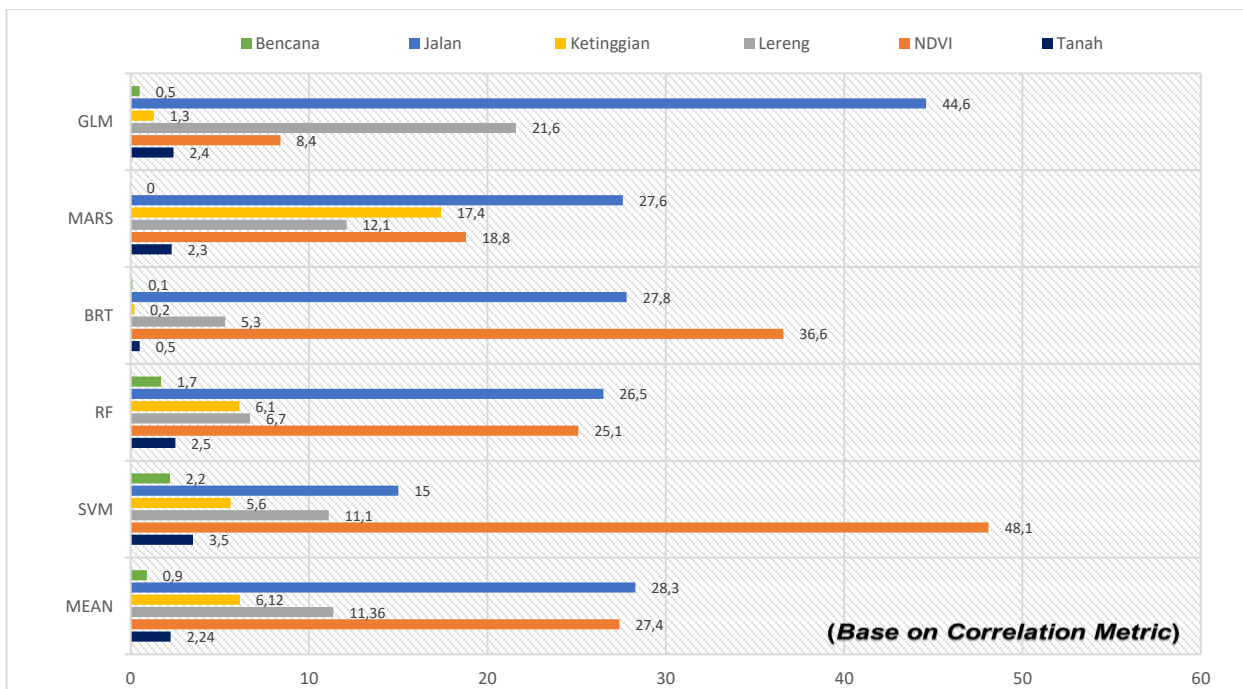
Dari model gabungan diperoleh bahwa pengaruh relatif rata-rata variabel jarak dari jalan menempati urutan pertama (28,3%), diikuti oleh variabel NDVI (27,4%). Hasil dari penelitian ini sejalan dengan penelitian yang dilakukan oleh Akpoti et al., (2019) yang menyatakan bahwa variabel jarak ke jalan merupakan faktor kritis dalam menentukan kelayakan ekonomi dan berpengaruh pada produktivitas padi. Selain itu, variabel indeks vegetasi (NDVI) juga memiliki pengaruh yang signifikan terhadap kesesuaian lahan untuk sawah (Nafi dan Basuki, 2019).

```
[1] "Bencana.tif" "Drainase" "Jalan.tif" "Ketinggian.tif" "Lereng.tif" "NDVI.tif" "Tanah.tif"
class      : RasterStack
dimensions : 8512, 9000, 76608000, 7 (nrow, ncol, ncell, nlayers)
resolution : 30, 30 (x, y)
extent     : 339645, 609645, 232085.5, 487445.5 (xmin, xmax, ymin, ymax)
crs       : +proj=utm +zone=50 +datum=WGS84 +units=m +no_defs
names     : Bencana, Drainase, Jalan, Ketinggian, Lereng, NDVI, Tanah
min values: 1, 1, 0, 1, 1, 0, 0
max values: 4.00000, 4.00000, 49.12642, 4.00000, 4.00000, 6.00000, 12.00000
```

Gambar 2. Hasil perbandingan data variabel lingkungan

```
1 variables from the 7 input variables have collinearity problem:
Drainase
After excluding the collinear variables, the linear correlation coefficients ranges between:
min correlation ( NDVI ~ Jalan ): 0.07623531
max correlation ( Lereng ~ Bencana ): -0.7314121
----- VIFs of the remained variables -----
Variables      VIF
1  Bencana 2.902050
2   Jalan 1.482787
3 Ketinggian 2.659672
4   Lereng 2.302688
5    NDVI 1.169116
6   Tanah 1.134983
```

Gambar 3. Hasil analisis VIF



Gambar 4. Pengaruh relatif variabel prediktor (%)

### 3.3. Hubungan data kehadiran spesies dan variabel lingkungan

Salah satu hasil analisis *species distribution modeling* (SDM) adalah memberikan gambaran hubungan dari data kehadiran spesies (*presence data*) dengan variabel lingkungan sebagai input data dalam proses analisis SDM yaitu berupa kurva respon (Gambar 5). Pada kurva respon, menunjukkan kurva semakin naik pada variabel SKL Bencana. Sebaran

penanaman padi dominan berada pada area kategori bencana rendah, terendah pada area kategori bencana tinggi.

Dalam hal variabel jarak dari jalan, kurva respons menggambarkan penurunan yang signifikan dari 0 km hingga 5 km, kemudian mengalami kenaikan pada jarak  $\geq 20$  km, di mana kurva tersebut datar. Oleh karena itu, dapat disimpulkan bahwa lokasi utama untuk menanam padi terletak dalam jarak yang dekat

dengan jalan, terutama di dalam kisaran 0 hingga 5 km. Selanjutnya, pada jarak  $\geq 20$  km dari jalan, terjadi peningkatan yang lambat dalam kurva respons, yang mengindikasikan bahwa di daerah dengan jarak tersebut, kemungkinan besar lokasi penanaman padi berdekatan dengan sungai (sebagai alternatif transportasi) atau berada di daerah lembah.

Ketika kita mempertimbangkan kesesuaian lahan sebagai lahan pertanian pangan berkelanjutan (LP2B), hal ini menunjukkan bahwa lahan pertanian yang lebih dekat dengan jalan lebih diinginkan, karena kedekatan dengan jalan dapat meningkatkan aktivitas pertanian (Andriawan et al., 2020). Justifikasi ini juga sejalan dengan temuan Tarigan dan Syumanjaya (2013), yang mengindikasikan bahwa peningkatan dalam kualitas infrastruktur dapat mengurangi biaya transportasi bagi petani dan mendorong lebih banyak petani untuk menjual hasil pertanian secara langsung ke pasar. Semakin dekat lahan pertanian dengan jalan, semakin baik aksesnya untuk mendukung kegiatan pertanian, sehingga lahan tersebut sangat cocok untuk LP2B.

Kurva respon pada variabel ketinggian terlihat puncak kurva berada pada kelas ketinggian 500 – 1000 mdpl dan pada ketinggian 0 – 100 mdpl. Hal tersebut menunjukkan bahwa karakter penanaman padi di Provinsi Kalimantan Utara dominan berada pada kedua kelas ketinggian tersebut. Sejalan dengan hal ini, Paski et al., (2017) menyebutkan tanaman padi dapat tumbuh pada ketinggian 0-1500 mdpl.

Pada variabel lereng (kemiringan lahan), kurva respon menunjukkan karakteristik lahan sebaran penanaman padi, dominan berada pada kelas kemiringan lahan 0 – 2% kemudian menurun sampai dengan kelas kemiringan lahan  $> 15$  %. Bobihoe et al., (2015) menyebutkan bahwa pada kemiringan lahan 0 – 2 % sangat baik diusahakan untuk tanaman pertanian, sedangkan kemiringan lahan sampai dengan 15 % masih dapat diusahakan untuk tanaman pertanian tetapi dengan terasering disertai penanaman pohon mengikuti garis kontur sebagai upaya mencegah longsor.

Variabel indeks vegetasi (NDVI) menunjukkan sebaran penanaman padi pada kelas indeks vegetasi (tanah terbuka/tubuh air) dikarenakan pada masa panen padi atau mulai kembali menanam padi. Puncak tertinggi sebaran penanaman padi pada kelas indeks vegetasi rendah yang sesuai dengan vegetasi jenis pertanian. Kurva menurun pada indeks vegetasi sedang dan tinggi yang merupakan dominan vegetasi kebun/perkebunan dan pepohonan.

Pada variabel tanah, terlihat kecenderungan peningkatan distribusi dari jenis tanah Podsol, dengan nilai tertinggi terdapat pada jenis tanah Glei. Temuan ini sejalan dengan informasi yang disajikan oleh Soeprahardjo dan Suhardjo (1978), yang memperkirakan bahwa sekitar 70% dari tanah sawah di dataran rendah di Indonesia masuk ke dalam ordo Inceptisol, Entisol, dan Vertisol (yang setara dengan Aluvial, tanah Glei, Regosol, dan Grumusol). Sekitar 22% dari tanah tersebut terdapat di lahan sawah

"uplands" di daerah vulkanik, dan masuk ke dalam ordo Ultisol, Inceptisol, Andisol, dan Alfisol (yang setara dengan Latosol, Regosol, Andosol, dan Mediteran). Sementara sekitar 6% ditemukan di lahan sawah yang memiliki tanah masam, termasuk dalam ordo Ultisol dan Oxisol (Podsolik Merah Kuning). Kondisi aktual lahan sawah di dataran rendah, sebagian besar (55%) didominasi oleh subordo Aquept dan Aquent (Aluvial dan Tanah Glei), terutama di daerah rawa atau yang memiliki air tanah dangkal. Di sisi lain, pada lahan sawah di daerah "uplands," lebih mendominasi (17%) oleh subordo Udept (Latosol dan Regosol), yang umumnya awalnya adalah tanah kering yang diubah menjadi lahan sawah.

#### 3.4. Distribusi spesies padi (sawah)

Kehadiran distribusi spesies padi (sawah) untuk model RF, GLM, BRT, MARS, RF, SVM dan *Ensemble* masing-masing adalah 6,34 %, 6,05%, 5,88%, 5,86%, 4,36% dan 8,66% dari luas daratan Provinsi Kalimantan Utara (Gambar 6). Selain itu, model *ensemble* digunakan untuk menghasilkan peta yang menunjukkan kesesuaian ekologis padi (sawah) pada berbagai tingkat distribusi. Ambang batas untuk model *ensemble* adalah 0,55 (Tabel 4), dan piksel di bawah ambang batas (*threshold*) dianggap sebagai "tidak diliputi" dan nilai piksel di atas ambang batas dibagi lagi menjadi tiga kelas sebagai "rendah," "sedang," dan "tinggi" dari kesesuaian spesies padi (sawah). Kesesuaian spesies padi (sawah) tiga kelas dimaksud masing-masing adalah 2,23%, 2,51% dan 3,93% dari luas daratan Provinsi Kalimantan Utara.

#### 3.5. Evaluasi Kinerja SDMS Untuk Prediksi Kesesuaian Distribusi Spesies Padi (Sawah)

Kinerja relatif SDMs untuk prediksi kesesuaian distribusi spesies padi (sawah) dengan melakukan penilaian evaluasi kinerja yang dari algoritma pembelajaran mesin (RF, BRT, dan SVM) dan model regresi (GLM dan MARS).

Penilaian kinerja model juga dapat memberikan dasar untuk membandingkan teknik pemodelan alternatif (Loiselle et al., 2003; Pearson et al., 2006) dan memungkinkan pengguna untuk menyelidiki bagaimana sifat yang berbeda dari data dan/atau spesies mempengaruhi keakuratan peta prediktif yang dihasilkan oleh model (Landmann et al., 2020; Reese et al., 2005; Seoane et al., 2005).

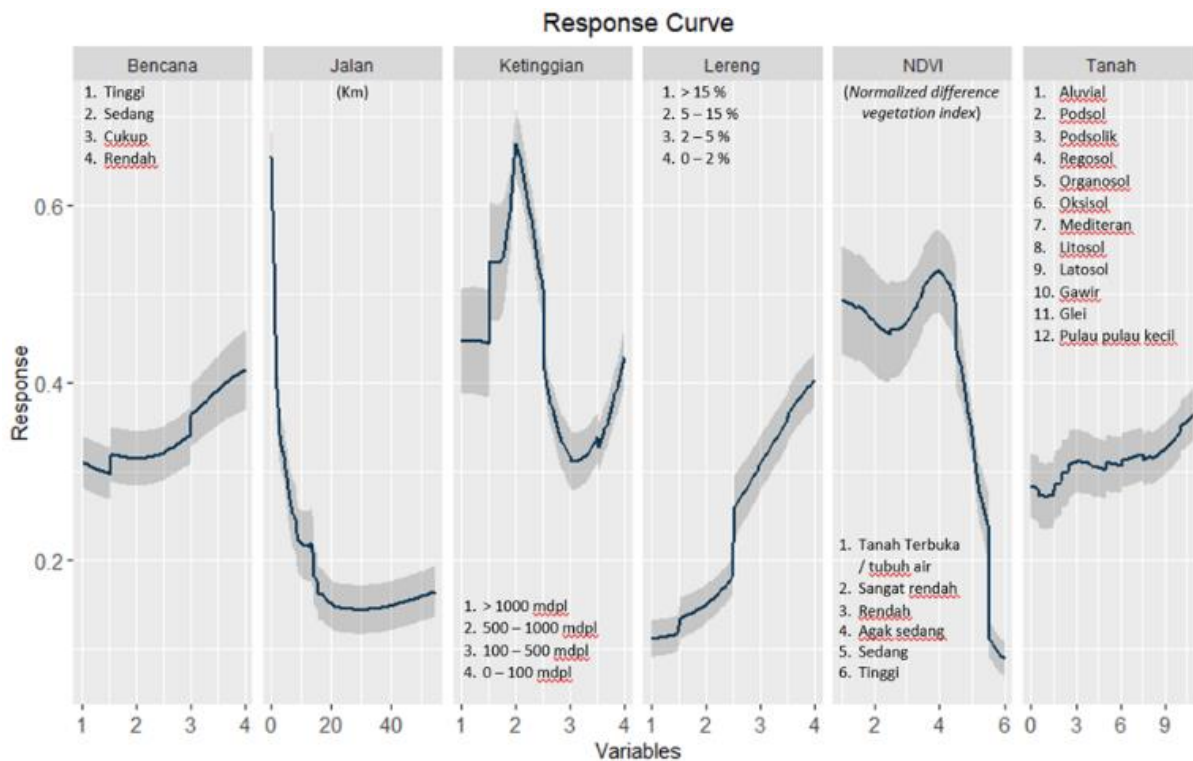
Menurut González-Ferreras et al., (2016), model dengan nilai AUC di bawah 0,7 dan di atas 0,9 masing-masing dianggap "sangat buruk" dan "sangat baik," Selain itu, model dengan nilai TSS dan kappa di bawah 0,4 dan di atas 0,8 masing-masing dianggap sebagai "sangat buruk" dan "sangat baik". Berdasarkan teknik evaluasi atas kinerja semua model, berada dalam kategori "baik sampai dengan sangat baik" sedangkan kinerja model GLM berkinerja buruk berdasarkan parameter *Cohen's Kappa* dan juga memiliki nilai hasil evaluasi kinerja

terendah dibandingkan kinerja model yang lainnya. Prediksi yang diperoleh dari model-model di atas, dibandingkan dengan penelitian sebelumnya, menunjukkan kinerja SDM memiliki implikasi besar dalam memberikan prediksi tertentu. Selain itu, banyak penelitian menyoroti kinerja RF yang lebih tinggi untuk prediksi berbasis penginderaan jauh dari distribusi spesies setelah mengevaluasi kinerja beberapa SDM (Jensen et al., 2020; West et al., 2016, 2017).

Area di bawah kurva ROC (AUC), kurva karakteristik operasi penerima (ROC) memberikan teknik alternatif untuk penilaian akurasi model skor ordinal (Fielding dan Bell, 1997). Konstruksi kurva ROC menggunakan semua ambang batas yang mungkin untuk mengklasifikasikan skor menjadi matriks kebingungan (*confusion matrix*), diperoleh nilai sensitivitas dan spesifisitas masing-masing matriks; kemudian membandingkan sensitivitas terhadap proporsi *false positif* yang sesuai (sama dengan 1 - spesifisitas). Menggunakan semua ambang batas (*threshold*) dengan menghindari pilihan

sewenang-wenang dari ambang batas tunggal (Liu et al., 2005; Manel et al., 2001), dan mempertimbangkan *trade-off* sensitivitas dan spesifisitas (Pearce dan Ferrier, 2000).

AUC terbukti independen dari prevalensi (Manel, Williams dan Ormerod, 2001; McPherson, Jetz dan Rogers, 2004) dan dianggap sebagai ukuran yang sangat efektif untuk kinerja model skor ordinal. Namun, aplikasi praktis model distribusi spesies dalam perencanaan konservasi, seperti identifikasi hotspot keanekaragaman hayati dan pemilihan lokasi konservasi yang representatif, seringkali memerlukan peta distribusi spesies yang menunjukkan ada dan/atau tidak ada, dan dengan demikian diperlukan pemilihan ambang batas untuk mengubah skor ordinal menjadi prediksi kehadiran-ketidakhadiran (Immitzer, et al., 2019; Howard, 2019). Dalam keadaan ini, akurasi evaluasi prediksi harus didasarkan pada ambang batas spesifik yang dipilih, sebagai lawan dari kurva ROC yang tidak bergantung pada ambang batas.

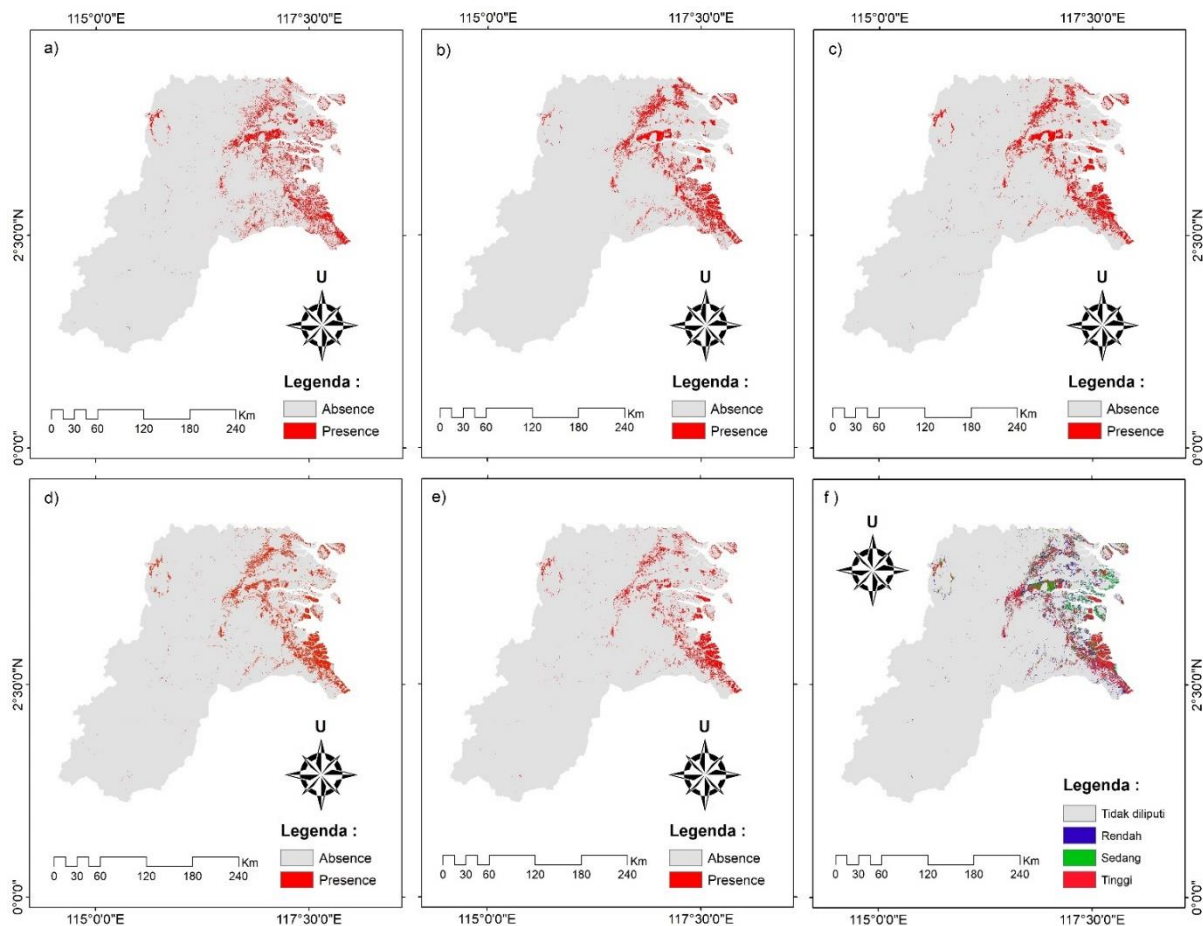


Gambar 5. Kurva respon data kehadiran spesies dengan variabel lingkungan

Tabel 4. Evaluasi kinerja SDM menggunakan parameter statistik yang berbeda

No	Metode	AUC	TSS	Kappa	COR	Sensitivity	Specificity	Threshold
1	GLM	0,86	0,61	0,57	0,65	0,82	0,79	0,73
2	MARS	0,94	0,78	0,76	0,82	0,91	0,87	0,72
3	BRT	0,92	0,73	0,70	0,75	0,89	0,85	0,76
4	RF	<b>0,96</b>	<b>0,80</b>	<b>0,79</b>	<b>0,83</b>	<b>0,93</b>	<b>0,88</b>	<b>0,67</b>
5	SVM	0,93	0,77	0,75	0,77	0,89	0,87	0,88
6	Ensamble	0,96	0,70	0,71	0,83	0,96	0,74	0,55





**Gambar 6.** Kehadiran distribusi padi (sawah)  
Keterangan : a) GLM; b) BRT; c) MARS; d) RF; e) SVM; f) Ensemble

Ukuran paling populer untuk keakuratan prediksi kehadiran – ketidakhadiran adalah kappa Cohen (Linders et al., 2020; Manel et al., 2001; Loiselle et al., 2003; Petit et al. 2003; Berg et al., 2004; Parra et al., 2004; Pearson et al., 2004; Rajah et al. 2018). Ukuran ini mengoreksi akurasi keseluruhan prediksi model dengan akurasi yang diharapkan terjadi secara kebetulan. Statistik kappa berkisar dari -1 hingga +1, di mana +1 menunjukkan kesepakatan sempurna dan nilai nol atau kurang menunjukkan kinerja yang tidak lebih baik dari acak (Cohen, 1960). Keuntungan lain dari kappa adalah kesederhanaannya, fakta bahwa kesalahan komisi dan kelalaian diperhitungkan dalam satu parameter, dan toleransi relatifnya terhadap nilai nol dalam matriks kebingungan (Manel et al., 2001).

Terlepas dari penggunaannya yang luas, beberapa penelitian telah mengkritik statistik kappa karena secara inheren bergantung pada prevalensi dan mengklaim bahwa ketergantungan ini memungkinkan bias dan artefak statistik pada perkiraan akurasi (Cicchetti dan Feinstein 1990; Lantz dan Nebenzahl 1996). Dalam sebuah studi yang berfokus pada evaluasi model distribusi spesies, McPherson et al., (2004) menggunakan simulasi numerik untuk menganalisis ketergantungan kappa pada prevalensi spesies yang dimodelkan dan menemukan bahwa kappa merespons variasi prevalensi secara unimodal. Berdasarkan temuan ini dapat disimpulkan bahwa sensitivitas kappa terhadap

prevalensi secara keseluruhan, oleh sebab itu membuat kappa tidak sesuai untuk perbandingan akurasi model spesies.

Allouche et al., (2006) selain memberikan bukti teoritis dan empiris bahwa kappa, adalah salah satu ukuran kinerja model yang paling banyak digunakan dalam ekologi, tetapi memiliki keterbatasan serius yang membuatnya tidak cocok untuk aplikasi tersebut. Sebagai alternatif disarankan menggunakan statistik keterampilan sejati (TSS). *True Skill Statistic* (TSS) dapat mengkompensasi kekurangan kappa dengan tetap menjaga semua kelebihanannya. Oleh karena itu, TSS sebagai ukuran sederhana dan intuitif untuk kinerja model distribusi spesies ketika prediksi dinyatakan sebagai peta kehadiran- ketidakhadiran. Sehingga model *ensemble* dibangun menggunakan metode rata-rata tertimbang (weighted mean) *True Skill Statistic* (TSS) dari semua model.

Hasil evaluasi model *ensemble* memperoleh akurasi yang dapat diterima dan nilainya mendekati akurasi dari model RF dengan nilai masing-masing 0,96; 0,70 dan 0,71 untuk AUC, TSS dan Kappa. Prediksi model *ensemble* mampu meningkatkan kinerja model dibandingkan model tunggal.

#### 4. Kesimpulan

Terdapat variasi antar-algoritma dalam respon variabel prediktor terhadap distribusi spesies padi (sawah) di wilayah Provinsi Kalimantan Utara.

Kehadiran distribusi spesies padi (sawah), berdasarkan pengaruh relatif variabel prediktor dipengaruhi oleh variabel jarak dari jalan (infrastruktur) dan variabel indeks vegetasi (NDVI). Prediksi kehadiran model distribusi spesies padi (sawah) pada model *ensemble* memiliki luasan terbesar dan model SVM memiliki luasan terkecil dibandingkan model lainnya. Tetapi hasil evaluasi kinerja model untuk hasil prediksi menunjukkan model RF merupakan model dengan kinerja lebih tinggi dibandingkan model lainnya berdasarkan parameter AUC, TSS, dan Kappa dan model *ensemble* memiliki kinerja mendekati dengan model RF. Model *ensemble* yang dibangun berbasis data dengan pendekatan pemodelan multi-algoritma memungkinkan untuk menilai variabilitas dalam kinerja algoritma. Hasil variasi antar-algoritma yang diamati, hasil peta prediksi *ensemble* menunjukkan kesesuaian distribusi spesies padi yang kokoh (*robust*) dan mampu meningkatkan kinerja model daripada yang dicapai oleh satu algoritma tunggal.

#### DAFTAR PUSTAKA

- Abdi, A. M., (2020), Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data, *GIScience & Remote Sensing*, 57(1), 1-20. doi: 10.1080/15481603.2019.1650447.
- Abdulhafedh, A., (2017), A Novel Hybrid Method for Measuring the Spatial Autocorrelation of Vehicular Crashes: Combining Moran's Index and Getis-Ord G<sub>i</sub>\* Statistic, *Open Journal of Civil Engineering*, 7(02), 208. doi: 10.4236/ojce.2017.72013.
- Ahmed, N., Atzberger, C., and Zewdie, W., (2020), Integration of remote sensing and bioclimatic data for prediction of invasive species distribution in data-poor regions: a review on challenges and opportunities, *Environmental Systems Research*, 9(1), 1-18. doi: [10.1186/s40068-020-00195-0](https://doi.org/10.1186/s40068-020-00195-0).
- Ahmed, N., Atzberger, C. and Zewdie, W., (2021). Species Distribution Modelling performance and its implication for Sentinel-2-based prediction of invasive *Prosopis juliflora* in lower Awash River basin, Ethiopia. *Ecol Process*, 10(1), 1-18. <https://doi.org/10.1186/s13717-021-00285-6>.
- Akpoti K., Kabo-Bah A.T., Dossou-Yovo E.R., Groen T.A., Zwart S.J., 2020. Mapping suitability for rice production in inland valley landscapes in Benin and Togo using environmental niche modeling. *Science of the total environment*. 709, 136165. doi:10.1016/j.scitotenv.2019.136165
- Allouche, O., Tsoar, A., and Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of applied ecology*, 43(6), 1223-1232. doi: 10.1111/j.1365-2664.2006.01214.x
- Andriawan, R., Martanto, R., and Muryono, S., (2020), Evaluasi Kesesuaian Potensi Lahan Pertanian Pangan Berkelanjutan Terhadap Rencana Tata Ruang Wilayah, *Tunas Agraria*, 3(3). doi: 10.31292/jta.v3i3.126.
- Aquino, D. D. N., Rocha Neto, O. C. D., Moreira, M. A., Teixeira, A. D. S., and Andrade, E. M. D., (2018), Use of remote sensing to identify areas at risk of degradation in the semi-arid region. *Revista Ciência Agrônômica*, 49, 420-429. doi: 10.5935/1806-6690.20180047
- Austin, M. P., (2002), Spatial prediction of species distribution: an interface between ecological theory and statistical modelling, *Ecological Modelling*, 157(2-3), 101-118. doi : 10.1016/S0304-3800(02)00205-3.
- Badan Pusat Statistik, (2019), *Konversi Gabah ke Beras Tahun 2018*, Badan Pusat Statistik, Jakarta. ISBN: 978-602-438-250-6.
- Badan Pusat Statistik Kalimantan Utara, 2020, *Provinsi Kalimantan Utara dalam Angka Tahun 2020*, Badan Pusat Statistik Kalimantan Utara, Tanjung Selor. ISSN: 2621-9891..
- Bobihoe, J., Asni, N., dan Endrizal, E., (2015), Kajian Teknologi Mina Padi di Rawa Lebak di Kabupaten Batanghari Provinsi Jambi, *Jurnal Lahan Suboptimal*, 4(1): 47-56.
- Byeon, D., Jung, S., Lee, and W.-H., (2018), Review of CLIMEX and MaxEnt for studying species distribution in South Korea. *Journal of Asia-Pacific Biodiversity*, 11(3), 325-333. doi:10.1016/j.japb.2018.06.002.
- Cicchetti, D. V., and Feinstein, A. R., (1990), High agreement but low kappa: II. Resolving the paradoxes, *Journal of clinical epidemiology*, 43(6), 551-558. doi: [0.1016/0895-4356\(90\)90159-m](https://doi.org/10.1016/0895-4356(90)90159-m).
- Cohen, J., (1960), A coefficient of agreement for nominal scales, *Educational and psychological measurement*, 20(1), 37-46. Doi: [10.1177/001316446002000104](https://doi.org/10.1177/001316446002000104).
- Cortes, C., and Vapnik, V., (1995), Support-vector networks. *Machine learning*, 20(3), 273-297. doi: 10.1109/64.163674.
- Chhogyel, N., Kumar, L., Bajgai, Y., and Jayasinghe, L. S., (2020), Prediction of Bhutan's ecological distribution of rice (*Oryza sativa* L.) under the impact of climate change through maximum entropy modelling, *The Journal of Agricultural Science*, 158(1-2), 25-37. doi: 10.1017/S0021859620000350.
- Dang, A. T. N., Kumar, L., and Reid, M., (2020), Modelling the Potential Impacts of Climate Change on Rice Cultivation in Mekong Delta, Vietnam., *Sustainability*, 12(22), 9608. doi:10.3390/su12229608.
- Davis, A. P., Gole, T. W., Baena, S., and Moat, J., (2012), The impact of climate change on indigenous arabica coffee (*Coffea arabica*): predicting future trends and identifying priorities, *PloS one*, 7(11), e47981. doi: 10.1371/journal.pone.0047981.
- Duan, J., and Zhou, G., (2012), Climatic suitability of double rice planting regions in China, *Scientia Agricultura Sinica*, 45(2), 218-227.
- Elith, J., and Leathwick, J., (2007), Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines, *Diversity and Distributions*, 13(3), 265-275. doi:10.1111/j.1472-4642.2007.00340.x
- Elith, J., Leathwick, J. R., and Hastie, T., (2008), A working guide to boosted regression trees, *Journal of animal ecology*, 77(4), 802-813. doi: 10.1111/j.13652656.2008.01390.x.
- Engler, R., Waser, L. T., Zimmermann, N. E., Schaub, M., Berdos, S., Ginzler, C., and Psomas, A., (2013), Combining ensemble modeling and remote sensing for mapping individual tree species at high spatial

- resolution. *Forest Ecology and Management*, 310, 64-73. doi: [10.4236/ojce.2017.72013](https://doi.org/10.4236/ojce.2017.72013).
- Fielding, A. H., and Bell, J. F., (1997), A review of methods for the assessment of prediction errors in conservation presence/absence models, *Environmental conservation*, 24(1), 38-49. doi: [10.1017/s0376892997000088](https://doi.org/10.1017/s0376892997000088).
- Friedman, J. H., (1991), Multivariate adaptive regression splines, *The annals of statistics*, 19(1), 1-67. doi: [10.1214/aos/1176347963](https://doi.org/10.1214/aos/1176347963).
- Früh, L., Kampen, H., Kerkow, A., Schaub, G. A., Walther, D., and Wieland, R., (2018), Modelling the potential distribution of an invasive mosquito species: comparative evaluation of four machine learning methods and their combinations, *Ecological Modelling*, 388, 136-144. doi: [10.1016/j.ecolmodel.2018.08.011](https://doi.org/10.1016/j.ecolmodel.2018.08.011).
- Fourcade, Y., Engler, J. O., Rödder, D., and Secondi, J., (2014), Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. *PLoS one*, 9(5), e9712. doi: [10.1371/journal.pone.0097122](https://doi.org/10.1371/journal.pone.0097122).
- Gu, H., Wang, J., Ma, L., Shang, Z., and Zhang, Q., (2019), Insights into the BRT (Boosted Regression Trees) method in the study of the climate-growth relationship of Masson pine in subtropical China, *Forests*, 10(3), 228. doi: [10.3390/f10030228](https://doi.org/10.3390/f10030228).
- Guisan, A., Edwards Jr, T. C., and Hastie, T., (2002), Generalized linear and generalized additive models in studies of species distributions: setting the scene, *Ecological modelling*, 157(2-3), 89-100. doi: [10.1016/s0304-3800\(02\)00204-1](https://doi.org/10.1016/s0304-3800(02)00204-1)
- Guisan, A., Zimmermann, N. E., Elith, J., Graham, C. H., Phillips, S., and Peterson, A. T., (2007), What matters for predicting the occurrences of trees: techniques, data, or species' characteristics?, *Ecological monographs*, 77(4), 615-630. doi: [10.1890/06-1060.1](https://doi.org/10.1890/06-1060.1).
- González-Ferreras, A. M., Barquín, J., and Peñas, F. J., (2016), Integration of habitat models to predict fish distributions in several watersheds of Northern Spain, *Journal of Applied Ichthyology*, 32(1), 204-216. doi: [10.1111/jai.13024](https://doi.org/10.1111/jai.13024).
- GRiSP.. (2013). *Rice Almanac: Source Book for One of the Most Important Activities on Earth*. Global Rice Science Partnership (GRiSP), International Rice Research Institute (IRRI), Los Banos, Philippines.
- Harrell Jr, F. E., Lee, K. L., and Mark, D. B., (1996), Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors, *Statistics in medicine*, 15(4), 361-387. doi: [10.1002/0470023678.ch2b\(i\)](https://doi.org/10.1002/0470023678.ch2b(i)).
- He, Q., and Zhou, G., (2016), Climate-associated distribution of summer maize in China from 1961 to 2010, *Agriculture, Ecosystems & Environment*, 232, 326-335. doi: [10.1016/j.agee.2016.08.020](https://doi.org/10.1016/j.agee.2016.08.020).
- Hijmans RJ, Elith J., (2019) Spatial distribution models. Di akses dari <https://rspsatial.org/sdm/SDM.pdf> tanggal 23 Mei 2022.
- Hossell, J. E., Ellis, N. E., Harley, M. J., and Hepburn, I. R., (2003), Climate change and nature conservation: Implications for policy and practice in Britain and Ireland, *Journal for Nature Conservation*, 11(1), 67-73. doi : [10.1078/1617-1381-00034](https://doi.org/10.1078/1617-1381-00034).
- Howard PL., (2019), Human adaptation to invasive species: a conceptual framework based on a case study meta-synthesis. *Ambio* 48:1401-1430. <https://doi.org/10.1007/s13280-019-01297-5>.
- Immitzer M, Neuwirth M, Böck S, Brenner H, Vuolo F, Atzberger C., (2019), Optimal input features for tree species classification in Central Europe Based on multi-temporal Sentinel-2 data. *Remote Sens* 11:2599. <https://doi.org/10.3390/rs11222599>.
- Jalaeian, M., Golizadeh, A., Sarafrazi, A., and Naimi, B., (2018), Inferring climatic controls of rice stem borers' spatial distributions using maximum entropy modelling, *Journal of Applied Entomology*, 142(4), 388-396. doi: [10.1111/jen.12493](https://doi.org/10.1111/jen.12493).
- Jensen, T., Seerup Hass, F., Seam Akbar, M., Holm Petersen, P., and Jokar Arsanjani, J., (2020), Employing machine learning for detection of invasive species using sentinel-2 and aviris data: The case of Kudzu in the United States, *Sustainability*, 12(9), 3544. doi: [10.3390/SU12093544](https://doi.org/10.3390/SU12093544)
- Kariyawasam, C. S., Kumar, L., and Ratnayake, S. S., (2019), Invasive plant species establishment and range dynamics in Sri Lanka under climate change. *Entropy*, 21(6), 571. doi: [10.3390/e21060571](https://doi.org/10.3390/e21060571).
- Kementerian Negara Lingkungan Hidup, (2009), *Peraturan Menteri Negara Lingkungan Hidup Nomor 17 Tahun 2009 tentang Pedoman Penentuan Daya Dukung Lingkungan Hidup Dalam Penataan Ruang Wilayah*, Kementerian Negara Lingkungan Hidup, Jakarta.
- Kementerian Pekerjaan Umum, (2007), *Peraturan Menteri Pekerjaan Umum Nomor 20/PRT/M/2007 tentang Teknik Analisis Aspek Fisik dan Lingkungan, Ekonomi, Serta Sosial Budaya dalam Penyusunan Rencana Tata Ruang*, Kementerian Pekerjaan Umum, Jakarta.
- Kosicki, J. Z., (2020), Generalised Additive Models and Random Forest Approach as effective methods for predictive species density and functional species richness, *Environmental and ecological statistics*, 27(2), 273-292. doi: [10.1007/s10651-020-00445-5](https://doi.org/10.1007/s10651-020-00445-5).
- Kogo, B. K., Kumar, L., Koech, R., and Kariyawasam, C. S., (2019), Modelling climate suitability for rainfed Maize cultivation in Kenya using a Maximum Entropy (MaxENT) approach, *Agronomy*, 9(11), 727. doi: [10.3390/agronomy9110727](https://doi.org/10.3390/agronomy9110727).
- Lantz, C. A., and Nebenzahl, E., (1996), Behavior and interpretation of the  $\kappa$  statistic: Resolution of the two paradoxes, *Journal of clinical epidemiology*, 49(4), 431-434. doi: [10.1016/0895-4356\(95\)00571-4](https://doi.org/10.1016/0895-4356(95)00571-4).
- Landmann T, Dubovyk O, Ghazaryan G, Kimani J, Abdel-Rahman E., (2020), Wide area invasive species propagation mapping is possible using phenometric trends. *ISPRS J Photogrammetry Remote Sens* 159:1-12. <https://doi.org/10.1016/j.isprs.2019.10.016>.
- Lamsal, P., Kumar, L., Aryal, A., and Atreya, K., (2018), Invasive alien plant species dynamics in the Himalayan region under climate change, *Ambio*, 47(6), 697-710. doi: [10.1016/S0304-3800\(02\)00205-3](https://doi.org/10.1016/S0304-3800(02)00205-3).
- Layomi Jayasinghe, S., Kumar, L., and Sandamali, J., (2019), Assessment of potential land suitability for tea (*Camellia sinensis* (L.) O. Kuntze) in Sri Lanka using a GIS-based multi-criteria approach, *Agriculture*, 9(7), 148. Doi: [10.3390/agriculture9070148](https://doi.org/10.3390/agriculture9070148).

- Leathwick, J. R., Elith, J., and Hastie, T., (2006), Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions, *Ecological Modelling*, 199(2), 188–196. doi: 10.1016/j.ecolmodel.2006.05.022.
- Lima, T. A., Beuchle, R., Langner, A., Grecchi, R. C., Griess, V. C., and Achard, F., (2019), Comparing Sentinel-2 MSI and Landsat 8 OLI imagery for monitoring selective logging in the Brazilian Amazon, *Remote Sensing*, 11(8), 961. doi: 10.3390/rs11080961.
- Linders T, Bekele K, Schaffner U, Allan E, Alamirew T, Choge S., (2020), The impact of invasive species on social-ecological systems: relating supply and use of selected provisioning ecosystem services. *Ecosyst Serv* 41:101055. <https://doi.org/10.1016/j.ecoser.2019.101055>.
- Liu, C., Berry, P. M., Dawson, T. P., and Pearson, R. G., (2005), Selecting thresholds of occurrence in the prediction of species distributions, *Ecography*, 28(3), 385-393. doi: 10.1111/j.0906-7590.2005.03957.x.
- Manel, S., Dias, J. M., and Ormerod, S. J., (1999), Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan River bird, *Ecological modelling*, 120(2-3), 337-347. doi: [10.1016/s0304-3800\(99\)00113-1](https://doi.org/10.1016/s0304-3800(99)00113-1).
- Manel, S., Williams, H. C., and Ormerod, S. J., (2001), Evaluating presence-absence models in ecology: the need to account for prevalence, *Journal of applied Ecology*, 38(5), 921-931. doi: [10.1046/j.1365-2664.2001.00647.x](https://doi.org/10.1046/j.1365-2664.2001.00647.x).
- McPherson, J. M., Jetz, W., and Rogers, D. J., (2004), The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact?, *Journal of applied ecology*, 41(5), 811-823. doi: [10.1111/j.0021-8901.2004.00943.x](https://doi.org/10.1111/j.0021-8901.2004.00943.x).
- Mi, C., Huettmann, F., Guo, Y., Han, X., and Wen, L., (2017), Why choose Random Forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence, *PeerJ*, 5, e2849. doi: 10.7717/peerj.2849.
- Mugo, R., and Saitoh, S.-I., (2020), Ensemble Modelling of Skipjack Tuna (*Katsuwonus pelamis*) Habitats in the Western North Pacific Using Satellite Remotely Sensed Data; a Comparative Analysis Using Machine-Learning Models. *Remote Sensing*, 12(16), 2591. doi:10.3390/rs12162591.
- Muthayya, S., Sugimoto, J.D., Montgomery, S. and Maberly, G.F., (2014), An overview of global rice production, supply, trade, and consumption, *Annals of the New York Academy of Sciences*, Vol. 1324, pp 7–14. doi: 10.1111/nyas.12540.
- Nafi A. Y., dan Basuki Y., 2019, Penentuan kawasan sawah berkelanjutan. *Jurnal Pembangunan Wilayah dan Kota*, 15(3), 214-226. doi: 10.14710/pwk.v15i3.21570
- Naimi, B., and Araújo, M. B., (2016), SDM: a reproducible and extensible R platform for species distribution modelling. *Ecography*, 39(4), 368-375. doi: 10.1111/ecog.01881.
- Ng, W. T., Meroni, M., Immitzer, M., Böck, S., Leonardi, U., Rembold, F., Gadain, H., & Atzberger, C., (2016), Mapping *Prosopis* spp. with Landsat 8 data in arid environments: Evaluating effectiveness of different methods and temporal imagery selection for Hargeisa, Somaliland, *Int J Appl Earth Observ Geoinf*, 53, 76-89. doi: 10.1016/j.jag.2016.07.019.
- Ng, W. T., Cândido de Oliveira Silva, A., Rima, P., Atzberger, C., and Immitzer, M., (2018), Ensemble approach for potential habitat mapping of invasive *Prosopis* spp. in Turkana, Kenya. *Ecology and evolution*, 8(23), 11921-11931. doi: 10.1002/ece3.4649.
- Parra, J. L., Graham, C. C., and Freile, J. F., (2004), Evaluating alternative data sets for ecological niche models of birds in the Andes, *Ecography*, 27(3), 350-360. doi: [10.1111/j.0906-7590.2004.03822.x](https://doi.org/10.1111/j.0906-7590.2004.03822.x).
- Paski, J. A. I., Faski, G. I. S. L., Handoyo, M. F., dan Pertiwi, D. S., (2017), Analisis neraca air lahan untuk tanaman padi dan jagung di Kota Bengkulu, *Jurnal Ilmu Lingkungan*, 15(2), 83-89. doi: [10.14710/jil.15.2.83-89](https://doi.org/10.14710/jil.15.2.83-89).
- Pearce, J., and Ferrier, S., (2000), Evaluating the predictive performance of habitat models developed using logistic regression, *Ecological modelling*, 133(3), 225-245. doi: [10.1016/s0304-3800\(00\)00322-7](https://doi.org/10.1016/s0304-3800(00)00322-7).
- Pearson, R. G., Dawson, T. P., and Liu, C., (2004), Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data, *Ecography*, 27(3), 285-298. Doi: [10.1111/j.0906-7590.2004.03740.x](https://doi.org/10.1111/j.0906-7590.2004.03740.x).
- Pearson, R. G., Thuiller, W., Araújo, M. B., Martinez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, T.E., and Lees, D. C., (2006), Model-based uncertainty in species range prediction, *Journal of biogeography*, 33(10), 1704-1711. doi: [10.1111/j.1365-2699.2006.01460.x](https://doi.org/10.1111/j.1365-2699.2006.01460.x).
- Petit, S., Chamberlain, D., Haysom, K., Pywell, R., Vickery, J., Warman, L., Allen, D., and Firbank, L., (2003), Knowledge-based models for predicting species occurrence in arable conditions, *Ecography*, 26(5), 626-640. doi: [10.1034/j.1600-0587.2003.03545.x](https://doi.org/10.1034/j.1600-0587.2003.03545.x).
- Raczko, E, and BogdanZ., (2018), Tree Species Classification of the UNESCO Man and the Biosphere Karkonoski National Park (Poland) Using Artificial Neural Networks and APEX Hyperspectral Images" *Remote Sensing* 10, no. 7: 1111. <https://doi.org/10.3390/rs10071111>.
- Rajah, P., Odindi, J., and Mutanga, O., (2018), Evaluating the potential of freely available multispectral remotely sensed imagery in mapping American bramble (*Rubus cuneifolius*). *South African Geographical Journal*, 100(3), 291-307. <https://doi.org/10.1080/03736245.2018.1461683>.
- Ratnayake, S. S., Kumar, L., and Kariyawasam, C. S., (2019), Neglected and underutilized fruit species in Sri Lanka: prioritisation and understanding the potential distribution under climate change, *Agronomy*, 10(1), 34. doi: 10.3390/agronomy10010034.
- Ravindra K, Rattan P, Mor S, and Aggarwal AN (2019), Generalized additive models: building evidence of air pollution, climate change and human health. *Environ Int* 132:104987. <https://doi.org/10.1016/j.envint.2019.104987>.
- Reese, G. C., Wilson, K. R., Hoeting, J. A., and Flather, C. H., (2005), Factors affecting species distribution predictions: a simulation modeling experiment, *Ecological Applications*, 15(2), 554-564. doi: [10.1890/03-5374](https://doi.org/10.1890/03-5374).
- Remya, K., Ramachandran, A., and Jayakumar, A. S., (2015), Predicting the current and future suitable habitat

Kusumadiya, P., Rusdiana, O., dan Mulatsih, S., (2024). Pemodelan Ensemble Prediksi Distribusi Ekologis Padi (*Oryza sativa*) di Provinsi Kalimantan Utara. *Jurnal Ilmu Lingkungan*, 22(2), 313-325, doi:10.14710/jil.22.2.313-325

- distribution of *Myristica dactyloides* Gaertn. using MaxEnt model in the Eastern Ghats, India. *Ecological engineering*, 82,184-188. doi: 10.1016/j.ecoleng.2015.04.053.
- roRustan, A. S., dan Aziza, T. N., (2008), Kompleksitas Penanganan Penguatan Ketahanan Pangan. *Jurnal Borneo Administrator*, 4(1). doi:10.24258/jba.v4i1.24.
- Seoane, J., Carrascal, L. M., Alonso, C. L., and Palomino, D., (2005), Species-specific traits associated to prediction errors in bird habitat suitability modelling, *Ecological Modelling*, 185(2-4), 299-308. doi: 10.1016/j.ecolmodel.2004.12.012.
- Soepraptohardjo, M., and Suhardjo, H., (1978), Rice soils of Indonesia, *Soils and rice*, 99-115.
- Soultan, A., and Safi, K., (2017), The interplay of various sources of noise on reliability of species distribution models hinges on ecological specialization, *PLoS one*, 12(11), e0187906. doi: 10.1371/journal.pone.0187906.
- Subroto, G., dan Susetyo, C., (2016), Identifikasi Variabel-Variabel yang Mempengaruhi Penentuan Lahan Pertanian Pangan Berkelanjutan di Kabupaten Jombang, Jawa Timur. *Jurnal Teknik ITS*, 5(2), C129-C133. doi: 10.12962/j23373539.v5i2.18297.
- Tarigan, S. D., dan Syumanjaya, R., (2013), Analisis Pengaruh Kualitas Infrastruktur Jalan terhadap Harga-Harga Hasil Pertanian di Kecamatan Dolok Silau, *Jurnal Ekonomi dan Keuangan*, 1(6), 14750.
- Thuiller, W., Guéguen, M., Renaud, J., Karger, D. N., and Zimmermann, N. E., (2019), Uncertainty in ensembles of global biodiversity scenarios, *Nature Communications*, 10(1), 1-9. doi: 10.1038/s41467-019-09519-w.
- Sabat-Tomala A, Edwin R, and Bogdan Z. (2020), Comparison of Support Vector Machine and Random Forest Algorithms for Invasive and Expansive Species Classification Using Airborne Hyperspectral Data" *Remote Sensing* 12, no. 3: 516. <https://doi.org/10.3390/rs12030516>.
- West, A. M., Evangelista, P. H., Jarnevich, C. S., Kumar, S., Swallow, A., Luizza, M. W., & Chignell, S. M., (2017), Using multi-date satellite imagery to monitor invasive grass species distribution in post-wildfire landscapes: An iterative, adaptable approach that employs open-source data and software. *Int J Appl Earth Observ Geoinf*, 59, 135-146. doi: 10.1016/j.jag.2017.03.009.
- West, A. M., Evangelista, P. H., Jarnevich, C. S., Young, N. E., Stohlgren, T. J., Talbert, C., Talbert, M., Morissette, J., and Anderson, R., (2016), Integrating remote sensing with species distribution models; mapping tamarisk invasions using the software for assisted habitat modeling (SAHM), *JoVE (Journal of Visualized Experiments)*, (116), e54578. doi: 10.3791/54578.
- Yoshida, S., (1981), *Fundamentals of Rice Crop Science*. International Rice Research Institute, Los Banos, Philippines.
- Zheng, B., Myint, S. W., Thenkabail, P. S., and Aggarwal, R. M., (2015), A support vector machine to identify irrigated crop types using time-series Landsat NDVI data. *International Journal of Applied Earth Observation and Geoinformation*, 34, 103-112. doi: 10.1016/j.jag.2014.07.002.