

**NASKAH PUBLIKASI
PENELITIAN**



***HOTSPOT PREDICTIVE MODELING USING REGRESSION
DECISION TREE ALGORITHM***

Oleh:

Dewi Asiah Shofiana, M. Kom.

Yohana Tri Utami, M. Kom.

Yunda Heningtyas, M. Kom.

Diterbitkan pada:

JURNAL TEKNOINFO

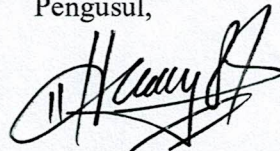
E-ISSN 2615-224X

**JURUSAN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
2022**

HALAMAN PENGESAHAN

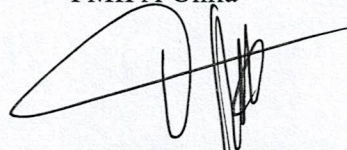
Nama : Yunda Heningtyas, M.Kom.
NIP : 19890108 201903 2 014
Jurusan : Ilmu Komputer
Fakultas : Matematika dan Ilmu Pengetahuan Alam
Judul Makalah : *Hotspot Predictive Modeling Using Regression Decision Tree Algorithm*
Oleh : Dewi Asiah Shofiana, Yohana Tri Utami, Yunda Heningtyas
Dimuat pada : Jurnal TEKNOINFO Vol. 16 No. 2 Juli 2022;
E-ISSN 2615-224X
Halaman 460-466
Diterbitkan oleh : Jurnal TEKNOINFO ©2022 Universitas Teknokrat Indonesia
Diunggah pada : <https://ejurnal.teknokrat.ac.id/index.php/teknoinfo/article/view/2051>

Bandar Lampung, 28 Oktober 2022
Pengusul,



Yunda Heningtyas, M. Kom.
NIP. 19890108 201903 2 014

Mengetahui,
Ketua Jurusan Ilmu Komputer
FMIPA Unila



Didik Kurniawan, S. Si, M.T.
NIP. 19800419 200501 1 004



Dr. Eng. Satripto Dwi Yuwono, M.T.
NIP. 19740705 200003 1 001

Menyetujui,
Ketua LPPM Universitas Lampung

Prof. Dr. Ir. Lusmeilia Afriani, D.E.A.
NIP.19650510 199303 2 008

SERTIFIKAT

Direktorat Jenderal Penguatan Riset dan Pengembangan,
Kementerian Riset, Teknologi, dan Pendidikan Tinggi



Kutipan dari Keputusan Direktur Jenderal Penguatan Riset dan Pengembangan,
Kementerian Riset, Teknologi, dan Pendidikan Tinggi Republik Indonesia
Nomor: 10/E/KPT/2019
Tentang Hasil Akreditasi Jurnal Ilmiah Periode 2 Tahun 2019

Jurnal TEKNOINFO

E-ISSN: 2615224X

Penerbit: Universitas Teknokrat Indonesia

Ditetapkan sebagai Jurnal Ilmiah

TERAKREDITASI PERINGKAT 4

Akreditasi berlaku selama 5 (lima) tahun, yaitu
Volume 13 Nomor 1 Tahun 2019 sampai Volume 18 Nomor 2 Tahun 2024

Jakarta, 4 April 2019

Direktur Jenderal Penguatan Riset dan Pengembangan



Dr. Muhammad Dimiyati
NIP. 195912171984021001





ISSN 2615-334X (online)
ISSN 1593-0010 (print)

JURNAL TEKNOINFO

Jurnal Teknik Informatika dan Ilmu Komputer
Terbit Enam bulanan (Januari dan Juli)

JURNAL

Penerbit **UNIVERSITAS TEKNOKRAT INDONESIA**

Pengelola **Program Studi S1 Informatika
Fakultas Teknik dan Ilmu Komputer**



VOL 16, NO 2 (2022)

JULI

DOI: <https://doi.org/10.33365/jti.v16i2>

FULL ISSUE

[View or download the full issue](#)

PDF

TABLE OF CONTENTS

ARTICLES

Analisis Sentimen Berbasis Aspek pada Wisata Halal dengan Metode Deep Learning

DOI : [10.33365/jti.v16i2.1516](https://doi.org/10.33365/jti.v16i2.1516)

Risca Naquitasia, Dthomas Hatta Fudholi, Lizda Iswari

PDF
156-164

EVALUASI RISIKO KEAMANAN INFORMASI DISKOMINFO PROVINSI XYZ MENGGUNAKAN INDEKS KAMI DAN ISO 27005 : 2011

DOI : [10.33365/jti.v16i2.1246](https://doi.org/10.33365/jti.v16i2.1246)

I Putu Setyo Syahindra, Clara Hetty Primasari, Aloysius Bagas Pradipta Iriantor

PDF
165-182

PEMBANGUNAN APLIKASI PENERJEMAH BAHASA ISYARAT DENGAN METODE CNN BERBASIS ANDROID

DOI : [10.33365/jti.v16i2.1752](https://doi.org/10.33365/jti.v16i2.1752)

Reza Haris Alfikri, Mardi Siswo Utomo, Hery Februriyanti, Eko Nurwahyudi

PDF
183-197

SISTEM PENDUKUNG KEPUTUSAN PENERIMAAN BEASISWA PENINGKATAN PRESTASI AKADEMIK MENGGUNAKAN METODE PREFERENCE SELECTION INDEX

DOI : [10.33365/jti.v16i2.1802](https://doi.org/10.33365/jti.v16i2.1802)

Rizal Tjut Adek, Hafizh Al Kautsar Aidilof, Ali Imran Nasution

PDF
198-212

Sistem Peringatan Dini Kebocoran Gas LPG Menggunakan ESP8266 dan API Telegram dengan Metode Fuzzy

DOI : [10.33365/jti.v16i2.1762](https://doi.org/10.33365/jti.v16i2.1762)

Dilan Allya Barqi, Mardi Siswo Utomo, Eddy Nurraharjo, Zuly Budiarmo

PDF
213-224

PENERAPAN ALGORITMA APRIORI PADA PENJUALAN SUKU CADANG KENDARAAN RODA DUA (Studi Kasus: Toko Prima Motor Sidomulyo)

DOI : [10.33365/jti.v16i2.1459](https://doi.org/10.33365/jti.v16i2.1459)

Zaenal Abidin, Arsya Kharisma Amartya, Arliyanti Nurdin

PDF
225-232

THE APPLICATION OF MACHINE LEARNING USING GOOGLE EARTH ENGINE FOR REMOTE SENSING ANALYSIS

DOI : [10.33365/jti.v16i2.1872](https://doi.org/10.33365/jti.v16i2.1872)

Muhammad Iqbal Habibie

PDF
233-237

APLIKASI PANDUAN GIZI MAKANAN BALITA BERBASIS ANDROID

DOI : [10.33365/jti.v16i2.1790](https://doi.org/10.33365/jti.v16i2.1790)

Yunita Rahma, Dini Suhartini, Sufiatul Maryana

PDF
238-245

- Focus and Scope
- Editorial Team
- Reviewer Team
- Author Guidelines
- Plagiarism Screening Policy
- Peer Review Process
- Online Submissions
- Publications Ethics
- Indexing
- Important Date
- Publication Fees
- Copyright & License
- Journal History
- Call for Reviewer

ACCREDITATION CERTIFICATE



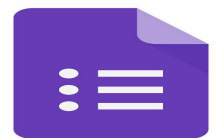
TEMPLATE IN INDONESIA



TEMPLATE IN ENGLISH



CALL FOR REVIEWER



Plagiarism Checker



Tools Reference



WEB SCRAPING OF DISEASE INFORMATION FROM SOCIAL MEDIA TWITTER

DOI : 10.33365/jti.v16i2.1871

Muhammad Iqbal Habibie, Taufiq Widiaputra, Yulianingsani Yulianingsani

PDF
246-251

A MULTICRITERIA INDEX USING NEURAL NETWORK TO EVALUATE THE POTENTIAL LANDS OF MAIZE

DOI : 10.33365/jti.v16i2.1873

Muhammad Iqbal Habibie, Nety Nurda

PDF
252-257

SISTEM ACCESS CONTROL RUANGAN LABORATORIUM DAN PERKULIAHAN MENGGUNAKAN RADIO FREQUENCY IDENTIFICATION

DOI : 10.33365/jti.v16i2.1897

Dian Kartika Utami, Akbar Sugih Miftahul Huda, Arie Qur'ania, Rizki Pratama

PDF
258-266

ANALISA KEGUNAAN DAN KEMUDAHAN LAYANAN MOBILE JKN PADA UPTD PUSKESMAS DEPOK DENGAN USE QUESTIONNAIRE DAN IPA

DOI : 10.33365/jti.v16i2.928

Rizal Nugraha, Lia Mazia, Lilyani Asri Utami

PDF
267-276

IMPLEMENTASI METODE RAPID APPLICATION DEVELOPMENT (RAD) DALAM RANCANGAN SISTEM INFORMASI PELAYANAN ADMINISTRASI DI KELURAHAN TABAM KOTA TERNATE

DOI : 10.33365/jti.v16i2.1889

M Kasyif Gufran Umar, Junaidi Sabtu, Rachmat Saleh Sukur

PDF
277-290

PEMANFAATAN FLUTTER PADA PENGEMBANGAN APLIKASI MOBILE EBISNIS PENYEDIAAN BAHAN BAKU BISNIS KATERING

DOI : 10.33365/jti.v16i2.1937

Arief Herdiansah, Desi Nurnaningsih, Hengki Rusdianto

PDF
291-303

E-CATALOG APPLICATION FOR FOOD AND BEVERAGES AT RUANG SEDUH CAFÉ BASED ON AUGMENTED REALITY

DOI : 10.33365/jti.v16i2.1957

Andi Nurkholis, Dyah Ayu Megawaty, M Fabian Apriando

PDF
304-312

MESIN PENTERJEMAH BAHASA INDONESIA-BAHASA SUNDA MENGGUNAKAN RECURRENT NEURAL NETWORKS

DOI : 10.33365/jti.v16i2.1930

Yustiana Fauziyah, Ridwan Ilyas, Fatan Kasyidi

PDF
313-322

PEMILIHAN PUPUK EFEKTIF UNTUK BUDIDAYA TANAMAN BAWANG MERAH DI KABUPATEN DEMAK

DOI : 10.33365/jti.v16i2.1931

Aditya Mahendra, Saefurrohman Saefurrohman

PDF
323-339

PENENTUAN MINAT KONSUMEN TERHADAP PRODUK MENGGUNAKAN ALGORITMA APRIORI PADA PT.TELKOM INDONESIA

DOI : 10.33365/jti.v16i2.1977

Pardo Frans Longgana, Irvan Irvan, Anjar Hero Wilarto

PDF
340-353

PENGEMBANGAN AUGMENTED REALITY MUSEUM CAKRANINGRAT BANGKALAN BERBASIS QR-CODE

DOI : 10.33365/jti.v16i2.1915

Prita Dellia, Siti Mutiatun, Ahmad Jami'ul Amil

PDF
354-363

IMPLEMENTASI DATA MINING TERHADAP DATA PENJUALAN DENGAN ALGORITMA APRIORI PADA PT. DUTA KENCANA SWAGUNA

DOI : 10.33365/jti.v16i2.1973

Yustina Meisella Kristania, Surya Listanto

PDF
364-372

SISTEM INFORMASI UMKM KAMPUNG GERABAH DESA PAGELARAN BERBASIS ANDROID

DOI : 10.33365/jti.v16i2.1998

Mohammad Nur Ali Furqon, Chaulina Alfianti Oktavia

PDF
373-383

IMPLEMENTASI METODE LOCATION BASED SERVICE PADA APLIKASI PENCARIAN

PDF

Current Indexing
DOAJ DIRECTORY OF
OPEN ACCESS
JOURNALS



More Abstracting & Indexing...



ISSN BARCODE

ISSN Online



ISSN Print

USER

Username

Password

Remember me

Login

LANGUAGE

Select Language

English

Submit

JOURNAL CONTENT

Search

Search Scope

All

Search

Browse

- » By Issue
- » By Author
- » By Title
- » Other Journals

FONT SIZE

| | |
|---|----------------|
| KOST DOI : 10.33365/jti.v16i2.2035 <i>Muhammad Bambang Firdaus, Ummul Hairah, Medi Taruk, Rosmasari Rosmasari, Mohammad Yasin, Lathifah Lathifah</i> | 384-394 |
| KLASIFIKASI KONSUMSI ENERGI INDUSTRI BAJA MENGGUNAKAN TEKNIK DATA MINING DOI : 10.33365/jti.v16i2.1984 <i>Sri Rahayu, Jajang Jaya Purnama</i> | PDF 395-407 |
| CITIZENAPP: YET ANOTHER SOCIAL MEDIA PLATFORM FOR PUBLIC FACILITY MONITORING IN SIDOARJO DOI : 10.33365/jti.v16i2.1985 <i>Jagad Yudha Awali, Irwan Alnarus Kautsar, Arif Senja Fitriani</i> | PDF 408-419 |
| APPLICATION OF AUGMENTED REALITY AS A LEARNING MEDIA OF MOLLUSCA GROUP ANIMAL RECOGNITION AND ITS HABITAT BASED ON ANDROID DOI : 10.33365/jti.v16i2.2044 <i>Olvi Lorensky Yeztiani, Qadhli Jafar Adrian, Ahmad Ari Aldino</i> | PDF 420-426 |
| DETEKSI RODA KENDARAAN DENGAN CIRCLE HOUGH TRANSFORM (CHT) DAN SUPPORT VECTOR MACHINE (SVM) DOI : 10.33365/jti.v16i2.1952 <i>Sri Dianing Asri, Desi Ramayanti, Ade Dwi Putra, Yohana Tri Utami</i> | PDF 427-434 |
| IMPLEMENTASI MULTI ATTRIBUTE DECISION MAKING MENGGUNAKAN METODE WEIGHTED PRODUCT DALAM PEMILIHAN SUPERVISOR DOI : 10.33365/jti.v16i2.2006 <i>Hidayanti Murtina, Nunung Hidayatun, Susafa'ati Susafa'ati</i> | PDF 435-442 |
| RANCANG BANGUN HELM PENDETEKSI DENYUT NADI DAN PEMBACA DOA PERJALANAN DOI : 10.33365/jti.v16i2.1989 <i>Shoffin Nahwa Utama, Abdul Wahid, Ahmad Fahmi Karami</i> | PDF 443-451 |
| ANALISIS TINGKAT PENERIMAAN APLIKASI BCA MOBILE DI KOTA MALANG MENGGUNAKAN METODE TECHNOLOGY ACCEPTANCE MODEL (TAM) DOI : 10.33365/jti.v16i2.2045 <i>Albert Christopher, Arif Tirtana, Addin Aditya</i> | PDF 452-459 |
| HOTSPOT PREDICTIVE MODELING USING REGRESSION DECISION TREE ALGORITHM DOI : 10.33365/jti.v16i2.2051 <i>Dewi Asiah Shofiana, Yohana Tri Utami, Yunda Heningtyas</i> | PDF 460-466 |
| SISTEM INFORMASI AKADEMIK PADA SMPIT TAMBUN ISLAMIC SCHOOL BEKASI DENGAN METODE WATERFALL DOI : 10.33365/jti.v16i2.2005 <i>Mely Mailasari, Monikka Nur Winnarto, Annida Purnamawati</i> | PDF 467-476 |

JURNAL TEKNOINFO

Published by Universitas Teknokrat Indonesia

Organized by Prodi S1 Informatika FTIK Universitas Teknokrat Indonesia

W: <http://ejurnal.teknokrat.ac.id/index.php/teknoinfo/index>

E: teknoinfo@teknokrat.ac.id

Jl. Zainal Abidin Pagaralam, No.9-11, Labuhan Ratu, Bandar Lampung



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Jumlah Pengunjung **100347694** View Teknoinfo StatsCounter

Visitors See more ▶

| | | | | | | | | |
|---|----|---------|---|----|-----|---|----|-----|
|  | ID | 110,286 |  | JP | 208 |  | CN | 117 |
|  | US | 4,417 |  | IN | 200 |  | NL | 108 |
|  | SG | 404 |  | RU | 146 |  | GB | 85 |
|  | MY | 372 |  | PH | 122 |  | TR | 81 |

Pageviews: 305,747





[Home](#) > [About the Journal](#) > **Editorial Team**

EDITORIAL TEAM

EDITOR-IN-CHIEF

Setiawansyah Setiawansyah, (Scopus ID : 57446859700), Universitas Teknokrat Indonesia, Indonesia

MANAGING EDITOR

Amarudin Amarudin, (Scopus ID: 57225264237, Universitas Teknokrat Indonesia), Indonesia

SECTION EDITORS

Nur Nawaningtyas Pusparini, (Scopus ID : 57222726782), STMIK Widuri, Indonesia

Agung Prajuhana Putra, (SCOPUS ID : 57211483540) Universitas Pakuan, Indonesia

Pritasari Palupiningsih, (Scopus ID : 57209461162) Teknik Informatika, Institut Teknologi PLN, Indonesia

Buce Trias Hanggara, (Scopus ID: 57200444099) Fakultas Ilmu Komputer, Universitas Brawijaya, Indonesia

Febry Eka Purwiantono, (ID Scopus : 57195732083) STIKI Malang, Indonesia

Fitri Retrialisca, (Scopus ID : 57213521397) Sistem Informasi, Universitas Airlangga, Indonesia

Arif Tirtana, (ORCID ID : 0000-0002-4289-3535) Sistem Informasi, STIKI Malang, Indonesia

Syaiful Ramdhan, (ID Scopus : 57215058900) ITB Bina Sarana Global, Indonesia

Purwono Prasetyawan, Institut Teknologi Sumatera , (Scopus ID : 57203960629)

Beny Prasetyo, [SCOPUS ID : 57195056275] Department of Information Systems, Jember University

Addin Aditya, (ID Scopus : 57205432625) STIKI Malang, Indonesia

Syaiful Ahdan, (SCOPUS ID: 57201820474, Universitas Teknokrat Indonesia), Indonesia

Sri Dianing Asri, (Sinta ID : 6664233) Universitas Dian Nusantara, Indonesia

Qadhli Jafar Adrian, (ID Scopus: 57210231478, Universitas Teknokrat Indonesia), Indonesia

Sufiatul Maryana, (Scopus ID : 57189051614) Sekolah Vokasi, Universitas Pakuan, Indonesia

Dedi Darwis, (Scopus ID : 57214720682) Universitas Teknokrat Indonesia, Indonesia

Arie Qur'ania, (Scopus ID : 57211206156) Ilmu Komputer, FMIPA, Universitas Pakuan, Indonesia

Yohana Tri Utami, (ORCID ID : 0000-0003-1082-1622) Ilmu Komputer, Universitas Lampung, Indonesia

Gibtha Fitri Laxmi, (Scopus ID : 55946516500), Universitas Ibn Khaldun Bogor, Indonesia, Indonesia

Wawan Gunawan, (Scopus ID : 55710884000) Universitas Islam Negeri Raden Intan Lampung, Indonesia

Adi Sucipto, (ID Scopus : 35325641900) Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia, Indonesia

Agus Ambarwari, (Scopus ID: 57193561828, Politeknik Negeri Lampung), Indonesia

Rizky Prabowo, (ID Scopus : 57220033327) Jurusan Ilmu Komputer, FMIPA, Universitas Lampung, Indonesia

Indonesia

LAYOUT EDITOR

Selamet Samsugi, FTIK, Universitas Teknokrat Indonesia

IT SUPPORTING/ADMINISTRATOR

Jupriyadi S.Kom., M.T., (Scopus ID : 56502396600 Universitas Teknokrat Indonesia) Indonesia

- » [Focus and Scope](#)
- » [Editorial Team](#)
- » [Reviewer Team](#)
- » [Author Guidelines](#)
- » [Plagiarism Screening Policy](#)
- » [Peer Review Process](#)
- » [Online Submissions](#)
- » [Publications Ethics](#)
- » [Indexing](#)
- » [Important Date](#)
- » [Publication Fees](#)
- » [Copyright & License](#)
- » [Journal History](#)
- » [Call for Reviewer](#)

ACCREDITATION CERTIFICATE



TEMPLATE IN INDONESIA



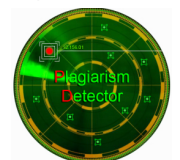
TEMPLATE IN ENGLISH



CALL FOR REVIEWER



Plagiarism Checker



Tools Reference





HOTSPOT PREDICTIVE MODELING USING REGRESSION DECISION TREE ALGORITHM

Dewi Asiah Shofiana¹⁾, Yohana Tri Utami²⁾, Yunda Heningtyas³⁾

^{1,2,3}Departement of Computer Science, Faculty of Mathematics and Natural Sciences, University of Lampung
^{1,2,3}Jl. Prof. Dr. Sumantri Brojonegoro No.1, Gedong Meneng, Rajabasa, Bandar Lampung, Lampung, Indonesia 35141
Email: ¹dewi.asiah@fmipa.unila.ac.id, ²yohana.utami@fmipa.unila.ac.id, ³yunda.heningtyas@fmipa.unila.ac.id

Abstract

Forest fires had always become an international issue influencing many life sectors, including environmental, social, and economic. The forest fire in 2013 was regarded as one of the worst forest fire tragedies in history, not only in Indonesia but also in the world. Detection of hotspots on the earth's surface by the satellite can be an indication of land and forest fire occurrence. This research aims to build a predictive model of monthly hotspots in Rokan Hilir Regency using the regression tree algorithm. Several variables related to weather information are included, such as rainfall, sea surface temperature, and southern oscillation index. This research used 245 training data and 43 testing data, resulting a predictive model with a correlation of 0.875 and an error rate of 0.166. Based on the values, we can conclude that the performance of the model is considerably good.

Keyword: forest fire, hotspot, regression tree.

1. INTRODUCING

Forest fire is a long-term problem in Indonesia that frequently become an international spotlight since it usually causes threats to various life sector, especially the environment, society, and economy. One indication of forest fires is the emergence of hotspots, in which the data is acquired daily by the satellite. Commonly, forest fire occurs due to two factors, namely human factors (land cover and land-use change, cultivation activities, and so on) or natural factors (weather and climate) [1]. Weather and climatic conditions such as rainfall, temperature, humidity, and wind speed affect the level of dryness of the earth's surface which can cause forest fires [2]. Moreover, most of the climate in Indonesia is related to the El Nino southern oscillation (ENSO) phenomenon, which climate variables especially rainfall is closely related to. El Nino and La Nina phenomenon will affect how much it rains on land. Depending on which cycle occurs, it can either cause droughts or flooding. Typically, El Nino is associated with drought, while La Nina is linked to increased flooding [3].

Riau Province is one region in Indonesia where forest fire frequently occurs since peatlands are still covering most of the areas. The phenomenal 2013 forest fire is considered one of the worst in Riau history, where the highest record is still held by the fire event in Riau between 1990-1997 [4]. According to the data released by Pekanbaru Meteorology Climatology and Geophysics Agency (BMKG) in August 2016, Rokan Hilir is the most fire-prone region, holding the highest distribution of hotspots with 46 hotspots data. On July 23, 2017, BMKG recapitulated the hotspot data in Riau Province, showing there were five hotspots detected in Rokan Hilir Regency with a confidence level of 50%. The higher the confidence level of hotspots, the stronger it indicates real fire existence [5].

[6] conducted a study regarding the prediction of hotspots emergence in Riau Province using the autoregressive integrated moving average (ARIMA) algorithm. The study shows that the monthly hotspot data obtained were not stationary in terms of variance and had a high value of mean absolute percentage error (MAPE) which is 40,974 due to the very high actual data in June, July, and August, resulting in a high margin of error. In 2015 [7] processed the same data but with the seasonal autoregressive integrated moving average (SARIMA) algorithm. The model from this study performs the best in predicting hotspots one month in the future, but it is required to be supported by the data of hotspot occurrences at least 13 months before. Furthermore, the inability of knowing the location of hotspots predicted is one shortcoming of the SARIMA algorithm. [8] used the Elman recurrent neural network (ERNN) as a method of making temporal predictions for the emergence of hotspots in Riau province. The model from this research has a MAPE value of 67.54% and a root mean square error (RMSE) of 252.98. But this ERNN model can only predict the possibility of hotspot emergence, without involving other variables that might highly affect the observed month of the study, such as rainfall.

Based on previous studies, this research builds a predictive model of monthly hotspots in the last 15 years using



climate data in Rokan Hilir Regency, Riau Province. Independent variables used are related to dry season rainfall, such as climate research unit (CRU) rainfall data, sea surface temperature (SST) NINO3.4, and southern oscillation index (SOI). The negative SOI value is closely related to the El Nino event which influences the prolonged dry season and drought in Indonesia. The more precise the predictor selected in the research, the better the resulting model [9]. Regression decision tree algorithm is implemented to process the data as we are expecting to predict continuous values of output (hotspots) based on independent variables. Besides, we are also working with a small amount of data which is suitable for this method.

2. RESEARCH METHODS

2.1 Data

Rokan Hilir Regency is selected as the area of interest of this research based on BMKG data that shows it as the most fire-prone area in Riau. This research utilizes hotspot data and climate data from 2001 to 2015 within Rokan Hilir area, which is specifically located at coordinates 11° 4' to 24° 5' North Latitude and 100° 17' to 101° 21' East Longitude. The hotspot datasets are collected from NASA Firms website (<https://firms.modaps.eosdis.nasa.gov>), the rainfall data are provided by the climatic research unit (<https://crudata.uea.ac.uk>), the global data of SOI are acquired from the Bureau of Meteorology (<http://www.bom.gov.au/climate/enso/soi/>), and SST Nino 3.4 data from the climate prediction center (<https://www.cpc.ncep.noaa.gov>). To build the model, this study applies the 'caret' and 'rpart' package available in RStudio.

2.2 Research Steps

This research was carried out in several stages: data acquisition, data preprocessing, data splitting with K-fold cross-validation, predictive modeling using regression tree, evaluation and model analysis.

2.2.1 Data Preprocessing

This step includes four main processes: data selection, forming the frequency matrix, creating a time series dataset, and data normalization.

a. Data Selection

Not all variables from the hotspot dataset are interesting for the research. The variables selected for further analysis are longitude, latitude, month of hotspots occurrences, and confidence.

b. Forming the Frequency Matrix

In determining the longitude and latitude range of Rokan Hilir area we refer to the rainfall dataset. The range is obtained from 9 grids on rainfall data that do not contain NA values (areas with no values detected by the satellite, commonly interpreted as the ocean area) [10]. Each latitude and longitude range is adjusted with hotspot data to calculate the monthly hotspot occurrence frequency. Selected Rokan Hilir areas for further analysis is the area within the research range with the highest total frequency of monthly hotspots emergences from 2001 to 2015 with a confidence level greater than 50%.

c. Create Time Series Dataset

Based on each month information for 15 years in the selected Rokan Hilir area, the selected data is adjusted with the other three variables data: SOI data, CRU rainfall, and SST Nino 3.4 data. We create a new time series dataset for the modeling and testing process. The dataset has independent variables (SOI, CRU, and SST) as well as response variables (the number of hotspots per month).

d. Data Normalization

Prior to the modeling process, we normalize the data specifically the data of the independent variables. This step creates the values of numeric columns in the dataset to use a common scale, without distorting differences in the ranges of values or losing information. Normalization of data is essential to avoid the dominance of values in data. Min-max normalization was carried out in this study, transforming the data values within the range of 0 to 1. According to [11], min-max normalization can be calculated using Equation 1.

$$\text{Norm}(x) = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$



With x is the data to be normalized, $\min(x)$ is the minimum data of the dataset, and $\max(x)$ refers to the maximum data.

2.2.2 Data Splitting

Data is split into two categories: training data and test data, using the k-fold cross-validation with $k = 8$. Cross-validation is a method to evaluate and compare algorithm performance by dividing the data into two parts, known as the part to train the model and the part to test the model. The k-fold cross-validation method begins by dividing the data into k data sets whose numbers are almost equal. The testing data for each iteration is the k -th dataset while the training data are the other datasets [12]. The percentage distribution with 8-fold cross-validation is 85% for training data and 15% for testing data from a total of 288 data. By this percentage, we can calculate that there are approximately 245 training data and 43 testing data utilized for modeling in each iteration. The data splitting process is conducted in R using the caret package.

2.2.3 Predictive Modeling with Regression Tree Algorithm

Predictive modeling of the monthly hotspots with regression tree algorithm is conducted in R using the rpart package. This process involves approximately 245 training data, where the resulted model is then tested using 43 testing data. The performance of the model is measured by the correlation value and error value. A higher correlation value and lower error value indicate a better model. The predicted results of the tested data are presented in a comparison graph between the number of occurrences of the actual monthly hotspots and the number of occurrences of the predicted monthly hotspots.

2.2.4 Evaluation and Model Analysis

The quality of model is measured by the value of the correlation coefficient (R) and the normalized root mean square error (NRMSE). Correlation coefficient shows the proximity of the relationship between actual and predicted response variables. The correlation value generally ranges from 1 to -1. Value closer to 1 or -1 indicate a stronger relationship between two variables. Conversely, value close to 0 shows that the relationship between the two variables is weak. Correlation can be calculated using Equation 2 [13].

$$R = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{[n \sum_{i=1}^n x_i^2 - [\sum_{i=1}^n x_i]^2][n \sum_{i=1}^n y_i^2 - [\sum_{i=1}^n y_i]^2]}} \quad (2)$$

With x_i shows actual monthly hotspots, shows y_i predicted monthly hotspots, and n shows the number of data row.

The error value is also calculated to measure the error rate in the predictive model built. Normalized root mean square error (NRMSE) is an alternative method to evaluate the accuracy of a predictive modeling process. NRMSE is the average value of the number of squares of errors, it can also state the size of the error generated by a predictive model. A low NRMSE value indicates that the variation in the value produced by a predictive model is close to the variation in the observed value. According to [14] NRMSE can be calculated using Equation 3.

$$\text{NRMSE} = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^n (x_i - y_i)^2}}{\sigma_y} \quad (3)$$

With x_i shows actual monthly hotspots, shows y_i predicted monthly hotspots, n shows the number of data row, and σ_y shows the standard deviation of the number of predicted hotspot occurrences. [15] stated that if the R value is close to 1 and the NRMSE value is close to 0, the model can be concluded as a good model.

3. RESULT AND DISCUSSIONS

3.1 Data Preprocessing

3.1.1 Forming the Frequency Matrix

The number of hotspots in Rokan Hilir before filtering it based on the confidence level are 48915 data. After filtering the data with confidence level greater and equal to 50%, the remaining number of hotspots reduced to 37295. Hotspot frequency matrix comes from the calculation of hotspots occurrences per month in an area. The area of interest is selected based on the number of hotspots. As a result, areas within the coordinates of 1.00 to



2.50 North Latitude and 100.00 to 101.50 East Longitude shows the highest frequency of hotspots occurrences in the 2000-2015 period, with a total of 25510 hotspots.

3.1.2. Create Time Series Dataset and Data Normalization

Dataset generated from this step has 34 variables: CRU1 to CRU27, SOI1 to SOI3, SST1 to SST3, and hotspots. There are 288 rows of data, in which each referring to hotspots data per month from January 2001 to December 2015. The variables SOI1 to SOI3 indicate SOI values taken from the Bureau of Meteorology site three months earlier. For example, the SOI1-SOI3 value in the January 2001 dataset is the SOI value from October to December 2000. This rule also applies to the CRU and SST data. Variables CRU1 to CRU27 indicate rainfall in the previous three months on the 9 grid areas selected in the CRU rainfall data. Variables SST1 to SST3 indicate SST values from the CPC site three months earlier. The hotspot variable is the response variable which implies the number of hotspots appearing per month. Variables SOI1 to SST3 are independent variables or predictor variables. The independent variables in the resulting dataset are then normalized. Calculation of data normalization uses min-max normalization as given in Equation 1.

3.2 Data Splitting

Data splitting in this study applies the 8-fold cross-validation, resulting training data and testing data with a ratio of 85:15 from 288 data. The training data consists of approximately 245 rows of data and approximately 43 testing data for each fold. The data splitting process implements the caret package available in R. The proportion of distribution of training data and testing data with 8-fold cross-validation is presented in Table 1.

Table 1. Distribution Proportion of Training Data and Testing Data for Each Fold

| Fold | Numbers of data (rows) | |
|------|------------------------|--------------|
| | Training data | Testing data |
| 1 | 245 | 43 |
| 2 | 247 | 41 |
| 3 | 246 | 42 |
| 4 | 245 | 43 |
| 5 | 244 | 44 |
| 6 | 246 | 42 |
| 7 | 247 | 41 |
| 8 | 245 | 43 |

3.3. Predictive Modeling with Regression Tree Algorithm

After normalizing the data on the independent variables, we build predictive models using the regression tree algorithm. Training data for each fold in Table 1 is used in developing the model, and normalization was also carried out on the actual response variable (Y) and the prediction results of the model. The model obtained was tested using testing data for each data fold as provided in Table 1. Figure 1 and Figure 2 show the best and worst graphs based on the testing results on all data folds.

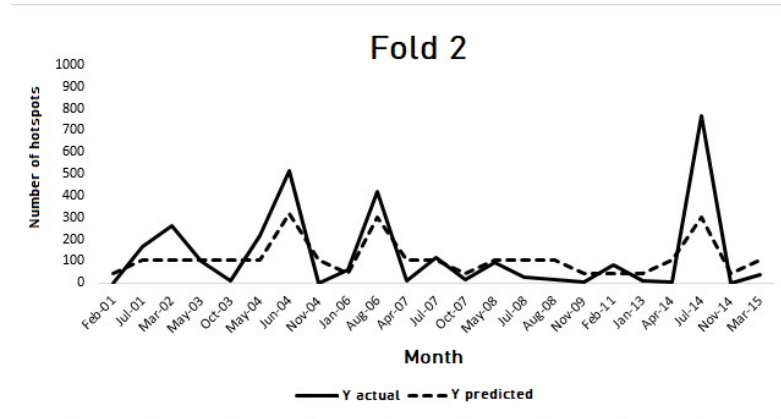


Figure 1. Best graph based on the testing results.

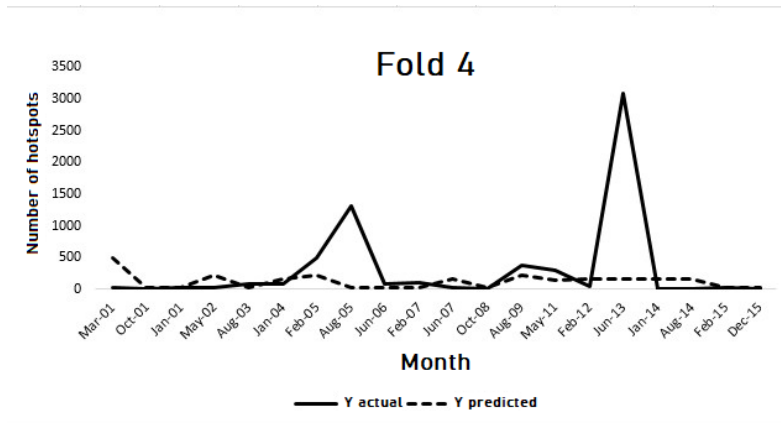


Figure 2. Worst graph based on the testing results.

The graph in Figure 1 is categorized as good since the two lines that represent the value of each actual variable and the predicted variable formed is not so different or aligned, while Figure 2 shows a contrary graph. In Figure 2 (4th fold) exist a very high value in the actual data while the predicted value is low, resulting the graphs of actual and predicted variables very different. As in Figure 1, the actual high value is the number of monthly hotspots detected high, which occurred in July 2014 with a total of 765 hotspots. The predicted value in the same month is 301 hotspots, making the comparison of the actual and predicted value quite far, but still considered as good since it does not prevail the alignment of the resulted graph. Figure 2 shows a comparison between the prediction and actual value in the 4th fold. A very high value in the actual variable occurred in July 2013 where 3097 hotspots are detected, while the predicted value at only 162 hotspots, causing extreme difference between both variables. Another extreme condition occurred in August 2005 with the actual value reaching 1313, but the predicted value was only 35. These events made the comparison graph out of alignment and considered as the worst graph compared to all folds. The incapability or error of the model in predicting can cause quite significant difference between the actual and the predicted. Prediction errors can occur due to several things, one of which is the possibility of including wet months (months with high rainfall) in the independent variables as predictors. A higher rainfall value can prevail the predicted value of the number of monthly hotspots do not match the actual value. The inability of the model in learning the pattern can also be a factor of prediction errors [16].

3.4. Evaluation and Model Analysis

Tested models are evaluated by calculating the correlation value and error value. We calculate the correlation is to see how strong the relationship between the response variable (Y) actual and predicted. Predictions of the number of monthly hotspots using a regression tree produces various correlation coefficients and NRMSE error values. Correlation and error values for each fold are presented in Table 2.



Table 2. Correlation and NRMSE value of each fold

| Fold | Correlation | NRMSE |
|---------|-------------|-------|
| 1 | 0.101 | 0.462 |
| 2 | 0.875 | 0.166 |
| 3 | 0.527 | 0.302 |
| 4 | 0.036 | 0.348 |
| 5 | 0.681 | 0.290 |
| 6 | 0.776 | 0.217 |
| 7 | 0.247 | 0.317 |
| 8 | 0.319 | 0.299 |
| Average | 0.445 | 0.301 |

A favored predictive model comes from the fold with a high value of correlation and a small value of NRMSE error. Table 1 shows that the 2nd fold has the best correlation and NRMSE results compared to others. The 2nd fold has a correlation value of 0.875 which is the highest correlation value of all, while the 4th fold has the smallest correlation value at only 0.036. The correlation value of the 2nd fold has a good predictive result of the number of monthly hotspots, as shown in Figure 1. On the other hand, the unreliable predictive model of the 4th fold can be seen in Figure 2. The extreme difference of actual and predicted values affecting the correlation coefficients to be low. In addition, the low correlation might also come from a bias toward data with small values.

The lowest NRMSE error value is 0.166 from the 2nd fold. The prediction error in the 2nd fold is very low, making the 2nd fold the best model based from both the correlation and NRMSE value. In addition, the 1st fold has the highest NRMSE compared to the others, reaching 0.462. Prediction error of monthly hotspots can affect the correlation, as seen in fold-1 with a correlation value at 0.101. The evaluation values of predictive models are influenced by the independent variables involved in the study. Table 3 shows the minimum, maximum, and average values of the independent variables of testing data involved in this study for each fold.

Table 3. Minimum, maximum, and mean values of the independent variables for each fold

| Fold | SOI | | | CRU | | | SST | | |
|------|--------|-------|-------|--------|--------|--------|-------|-------|-------|
| | Min | Max | Mean | Min | Max | Mean | Min | Max | Mean |
| 1 | -14.70 | 21.40 | 0.76 | 71.10 | 432.70 | 214.89 | 25.01 | 28.93 | 27.19 |
| 2 | -15.90 | 22.30 | -0.67 | 60.50 | 567.80 | 236.04 | 25.32 | 28.16 | 27.02 |
| 3 | -16.00 | 25.10 | 1.67 | 56.60 | 520.20 | 225.86 | 25.00 | 28.85 | 26.71 |
| 4 | -18.60 | 13.90 | -2.04 | 63.60 | 548.40 | 238.34 | 26.05 | 29.26 | 27.07 |
| 5 | -12.00 | 21.30 | 1.91 | 66.90 | 430.90 | 216.63 | 25.00 | 28.03 | 27.06 |
| 6 | -11.20 | 17.10 | -0.64 | 108.40 | 462.77 | 242.90 | 25.65 | 28.15 | 27.17 |
| 7 | -19.80 | 24.90 | 2.03 | 58.10 | 298.10 | 231.37 | 25.03 | 28.93 | 27.20 |
| 8 | -12.00 | 18.30 | 1.92 | 55.30 | 491.60 | 223.72 | 24.86 | 28.90 | 27.01 |

As we can see in Table 3, the 2nd fold testing data has a minimum SOI of -15.90, a maximum SOI of 22.30 and an average SOI of -0.67. Judging from the minimum negative value, it shows the El Nino phenomenon occurs, where this phenomenon is related to the long dry season and drought in the southern area, including Indonesia. In the second fold, there is a minimum CRU of 60.50, a maximum CRU of 567.80, and an average CRU of 236.04. This fold also has a minimum SST of 25.32, a maximum SST of 28.13, and an average SST of 27.02. Viewing the values of the climatic variables, the sea surface temperature from the 2nd fold predictive model is not too high as indicated by the maximum SST value, but it has the highest maximum rainfall value compared to the other folds. This situation appears since the 2nd fold testing data has various months, including the months both from earlier and later of the year, making the data vary since rainy and dry season all exist in this fold.



The 4th fold model is considered the worst, with the lowest correlation and a high error value. In this fold, the minimum SOI value is very low at -18.60, the maximum SOI is 13.90, and the average is -2.04. The maximum and average SOI value in this fold is considered the lowest among the others, but the CRU rainfall value and sea surface temperature are high. Low SOI values indicate El Nino phenomenon exist, but it is contrary to the CRU value that shows high rainfall intensity and sea surface temperatures. The data shows there is a discrepancy between the climatic data variables, which all affect the predicted value of the number of monthly hotspots. Therefore, the predictive model in the 4th fold cannot recognize the pattern of the test data properly.

4. CONCLUSION

This research successfully builds a predictive model using a regression tree algorithm. The model evaluation shows an average correlation score of 0.445 and the NRMSE error with an average of 0.301. Generally, the models developed in this study have a correlation value in the moderate range, while the error value is considerably small. The 2nd fold performs the best with a correlation value of 0.875 and an NRMSE value of 0.166. Based on the correlation and error values, the 2nd fold model can be considered a good predictive model in forecasting monthly hotspots in Rokan Hilir Regency, Riau Province. In this study, there are folds that have low correlations. The model resulted from this study can only predict the number of monthly hotspots without knowing the location of their occurrence. Therefore, spatial variables can be added in further research in order to predict the location of hotspots. It is also necessary to pay attention to other aspects related to the occurrence of hotspots and forest fires, such as temperature or humidity, so that the prediction accuracy is better and more evenly distributed for each fold. In addition, future research can also use hotspots with a confidence level of 70% since a higher confidence level indicates a higher possibility a fire indeed exists.

REFERENCES

- [1] N. Dina, "Identifikasi sumber api penyebab kebakaran, riam kanan kalimantan selatan," Banjarbaru, 2011. [Online]. Available: http://eprints.ulm.ac.id/187/1/Mandiri_Sumber_Api_Kebakaran.pdf.
- [2] L. Syaufina, *Kebakaran Hutan dan Lahan di Indonesia: Perilaku Api, Penyebab dan Dampak Kebakaran*. Malang: Bayumedia Publishing, 2008.
- [3] W. Estingtyas and A. H. Wigena, "Teknik Statistical Downscaling Dengan Regresi Komponen Utama Dan Regresi Kuadrat Terkecil Parsial Untuk Prediksi Curah Hujan Pada Kondisi El Nino, La Nina, Dan Normal," *J. Meteorol. dan Geofis.*, vol. 12, no. 1, pp. 65–72, 2011, doi: 10.31172/jmg.v12i1.87.
- [4] Kistyarini, "Pakar: Asap Riau Terparah Sepanjang Sejarah," *Kompas*, 2013.
- [5] D. A. Shofiana and I. S. Sitanggang, "Confidence Analysis of Hotspot as Peat Forest Fire Indicator," *J. Phys. Conf. Ser.*, vol. 1751, p. 012035, 2021, doi: 10.1088/1742-6596/1751/1/012035.
- [6] I. S. Robby, "Prediksi Temporal untuk Kemunculan Titik Panas di Provinsi Riau Menggunakan Autoregressive Integrated Moving Average (ARIMA)," Institut Pertanian Bogor, 2014.
- [7] N. Istiqamah, "Prediksi kemunculan titik panas di provinsi Riau menggunakan seasonal autoregressive integrated moving average (SARIMA)," Institut Pertanian Bogor, 2015.
- [8] T. T. Amaranggana, "Prediksi Temporal untuk Kemunculan Titik Panas di Provinsi Riau menggunakan Elman Recurrent Neural Network," Institut Pertanian Bogor, 2016.
- [9] A. Qadir, N. R. Talukdar, M. M. Uddin, F. Ahmad, and L. Goparaju, "Predicting forest fire using multispectral satellite measurements in Nepal," *Remote Sens. Appl. Soc. Environ.*, vol. 23, no. March, p. 100539, 2021, doi: 10.1016/j.rsase.2021.100539.
- [10] L. Giglio, W. Schroeder, and C. O. Justice, "The collection 6 MODIS active fire detection algorithm and fire products," *Remote Sens. Environ.*, vol. 178, pp. 31–41, 2016, doi: 10.1016/j.rse.2016.02.054.
- [11] Y. K. JAIN and S. K. BHANDARE, "Min Max Normalization Based Data Perturbation Method for Privacy Protection," *Int. J. Comput. Commun. Technol.*, vol. 4, no. 4, pp. 233–238, 2013, doi: 10.47893/ijct.2013.1201.
- [12] L. LIU and M. T. ÖZSU, *Encyclopedia of Database Systems*. Boston: Springer, 2009.
- [13] S. Siegel, *Statistic Non Parametrik Untuk Ilmu-Ilmu Sosial*. Jakarta: Gramedia Pustaka Utama, 1986.
- [14] D. C. Montgomery, C. L. Jennings, and M. Kulahci, *Introduction to Time Series Analysis and Forecasting, 2nd Edition*. New Jersey: John Wiley, 2015.
- [15] T. Wahyono, *Analisis Regresi dengan MS Excel 2007 dan SPSS 17*. Jakarta: PT Elex Media Komputindo, 2010.
- [16] I. Hakiki, A. Ihwan, and J. Sampurno, "Prediksi Kemunculan Titik Panas (Hotspot) Menggunakan Metode Jaringan Syaraf Tiruan Propagasi Balik Studi Kasus di Pontianak," *Prism. Fis.*, vol. 3, no. 2, pp. 75–78, 2015.