

## INTISARI

### MODEL REGRESI LOGISTIK TERBOBOTI GEOGRAFIS BERKORELASI TEMPORAL BERBASIS COPULA

Oleh

PEPI NOVIANTI

19/450318/SPA/00700

Regresi Terboboti Geografis (RTG) telah mendapatkan perhatian besar dalam analisis spasial karena memungkinkan estimasi parameter setiap lokasi dan penyajian dalam bentuk pemetaan. Pada data biner dengan efek spasial, model regresi yang dapat digunakan adalah Regresi Logistik Terboboti Geografis (RLTG). Banyak variabel menarik dalam geosains yang diamati tidak hanya lintas ruang, tetapi juga dipengaruhi efek waktu. Oleh karena itu, pengembangan model dengan informasi spasial dan temporal terus dikembangkan, di antaranya Regresi Terboboti Geografis dan Temporal (RTGT) dan Regresi Logistik Terboboti Geografis Berkorelasi Temporal (RLTGKT).

RLTGKT mengintegrasikan model RLTG dan pendekatan estimasi dua langkah pada data longitudinal. Untuk menghubungkan korelasi antara data spasiotemporal dengan struktur autokorelasi AR(1), RLTGKT mengadopsi konsep korelasi *tetrachoric*. RLTGKT mampu mengestimasi koefisien di setiap lokasi dan seluruh periode waktu, serta memvisualisasikan nilai dan pengaruh yang bervariasi secara spasial dan temporal. Namun, algoritma yang digunakan relatif besar dan membutuhkan waktu komputasi yang relatif lama. Selain itu dengan asumsi model marjinal berdistribusi Bernoulli, RLTGKT digunakan untuk memodelkan data spasiotemporal Binomial dan pembobot geografis dilakukan dengan pendekatan model Regresi Poisson Terboboti Geografis (RPTG).

Distribusi Binomial dan Poisson memiliki variabel acak yang berbeda, sehingga pemilihan *bandwidth* optimum dengan RPTG kurang sesuai untuk model RLTGKT. Oleh karena itu perlu digunakan model RLTG dengan asumsi data Binomial dalam penentuan *bandwidth* optimum. Untuk pengaruh temporal, korelasi *tetrachoric* digunakan untuk menentukan probabilitas gabungan dari dua variabel dependen bersifat biner dengan menggunakan probabilitas marjinal berupa probabilitas logistik. Selain korelasi *tetrachoric*, model copula juga dapat digunakan untuk menentukan probabilitas gabungan dari dua variabel dan mampu bersaing dengan metode probabilitas gabungan lainnya. Metode copula memiliki beberapa kelebihan

yaitu memiliki interpretasi yang lebih jelas dan lebih mudah untuk dikembangkan untuk respon bivariat dengan tiga atau lebih kategori.

Dengan mengasumsikan data spasiotemporal berupa data Binomial, model RLTGKT dapat dikembangkan pada model RLTGKT data Binomial. Keterbaruan dari model RLTGKT berbasis copula pada data Binomial ini adalah menerapkan model RLTGB untuk menentukan pembobot geografis dan menggunakan pendekatan berbasis copula dengan model Regresi Logistik Rantai Markov Berbasis Copula (RLRMBC) untuk korelasi waktu. Dengan demikian model RLTGKT yang diusulkan dalam penelitian ini adalah model RLTGKT berbasis copula pada data Binomial. Tujuan penelitian ini adalah menentukan *bandwidth* spasial menggunakan model RLTG data Binomial, menentukan korelasi waktu dengan dependensi copula dan menentukan estimasi parameter model RLTGKT berbasis copula pada data Binomial dan mengaplikasikannya pada data nyata.

Model RLTG data Binomial (RLTGB) dan estimasi koefisien parameter dapat menggambarkan keragaman hubungan spasial yang berkaitan dengan variabel independen dan variabel dependen. Berdasarkan hasil simulasi estimasi parameter RLTGB diperoleh hasil konsistensi yang relatif tinggi pada nilai parameter bangkitan dan parameter dugaan dengan korelasi Pearson untuk  $\beta_1$  adalah 0.989 ( $p < 0.0000$ ) dan  $\beta_2$  adalah 0.999 ( $p < 0.0000$ ). Untuk data deret waktu Binomial, model Rantai Markov Berbasis Copula (RMBC) dapat digunakan untuk memodelkan dependensi waktu. Model RMBC juga dapat memodelkan data deret waktu Binomial dengan variabel kovariat menggunakan model regresi logistik dan disebut Model Regresi Logistik Rantai Markov Berbasis Copula (RLRMBC). Model RLRMBC tidak hanya dapat melihat hubungan antara variabel independen dan dependen, tetapi juga memberikan estimasi dependensi antar waktu dari variabel dependen. Pada data spasiotemporal Binomial, model RLTGB dan RLRMBC dapat dikombinasikan untuk memodelkan ketakstasioneran variabel dependen akibat adanya efek spasial dan temporal, dengan demikian model ini juga dapat menangkap pengaruh spasial dan temporal pada data Binomial.

Model RLTGKT berbasis copula diaplikasikan pada data kasus Covid-19 di 34 provinsi di Indonesia pada periode waktu 1 Juli - 30 September 2021. Berdasarkan model RLTGKT berbasis copula diperoleh nilai  $R^2$  sebesar 90.92%, KTG sebesar 180,762 dan MAPE sebesar 63.10%. Dengan menggunakan Copula Clayton, Model RLTGKT dapat menangkap pengaruh spasial dan temporal yang berbeda-beda pada variabel faktor cuaca terhadap penyebaran Covid-19 di Indonesia.

**Kata-kata kunci:** Regresi Logistik, Distribusi Binomial, Spasial, Regresi Terboboti Geografis, Temporal, Dependensi Waktu, Copula Clayton, Copula Gumbel, Copula Frank.

## ***ABSTRACT***

### **COPULA-BASED GEOGRAPHICALLY WEIGHTED TEMPORALLY CORRELATED LOGISTIC REGRESSION MODEL**

By

PEPI NOVIANTI

19/450318/SPA/00700

Geographically Weighted Regression (GWR) has received great attention in spatial analysis because it allows the estimation of parameters for each location and presentation in the form of maps. In binary data with spatial effects, the regression model that can be used is Geographically Weighted Logistic Regression (GWLR). Many interesting variables in geoscience are observed not only across space but also affect the effects of time. Therefore extended models with spatial and temporal information continue to be developed, including Geographically and Temporal Weighted Regression (GWTR) and Geographically Weighted Temporally Correlated Logistic Regression (GWTCLR).

GWTCLR integrates the GWLR model and two-step estimation on longitudinal data. To relate the correlation between the spatiotemporal data and the AR(1) autocorrelation structure, GWTCLR adopts the concept of tetrachoric correlation. GWTCLR is able to estimate coefficients at each location and across time periods, and visualize the values and effects that vary temporally and spatially in a transparent and comprehensive way. However, the algorithm used is relatively large and requires a relatively long computation time. In addition, with the assumption that the marginal model has a Bernoulli distribution, GWTCLR is used to model Binomial spatiotemporal data and geographically weighted is carried out using a Geographically Weighted Poisson Regression (GWPR) model approach.

The Binomial and Poisson distributions are different random variables, accordingly choosing the optimal bandwidth with GWPR is not suitable. Therefore it is necessary to use the RLTG model with the assumption of Binomial data in determining the optimum bandwidth. For temporal effects, tetrachoric correlation is used to determine the joint probability of the two dependent variables which are binary by using marginal in the form of logistic probabilities. In addition to tetrachoric correlation, the copula model can also be used to determine the joint probabilities of two variables and is able to compete with other combined probability methods. The copula method has several advantages, namely having a clearer interpretation

and easier to develop for bivariate responses with three or more categories.

By assuming the spatiotemporal data is in the form of Binomial data, the GWTCLR model can be developed on the GWTCLR Binomial data model. The novelty of the copula-based GWTCLR model is to apply the GWLR model to determine geographically weighted and to use a copula-based approach with the Copula-based Markov Chain logistic regression (CBMCLR) model for time correlation. Therefore, the GWTCLR model proposed in this study is the copula-based GWTCLR model on Binomial data. The aims of this study are to determine the spatial bandwidth using the Binomial data GWLR model, to determine the time correlation with copula dependencies and to determine copula-based GWTCLR model parameter estimates on Binomial data and to apply the model to real data.

The GWLR model for Binomial data (GWLBR) and parameter coefficient estimation can describe the diversity of spatial relationships associated with the independent and dependent variables. Based on the GWLBR parameter estimation simulation results, relatively high consistency results were obtained for the generated parameter values and estimated parameters with the Pearson correlation for  $\beta_1$  being 0.989 ( $p < 0.0000$ ) and  $\beta_2$  is 0.999 ( $p < 0.0000$ ). For Binomial time series data, the Copula-based Markov Chain (CBMC) model can be used to model the time dependency. Using the CBMC Logistic Regression (CBMCLR) Model can also be used in modelling Binomial time series data with covariate variables. The CBMCLR model can not only see the relationship between the independent and dependent variables but also provide estimates of the intertemporal dependence of the dependent variable. In spatiotemporal Binomial data, the GWLBR and CBMCLR models can be combined to model the stationarity of the dependent variable due to spatial and temporal effects. This copula-based GWTCLR model can also capture spatial and temporal influences on Binomial data.

The copula-based GWTCLR is applied to data on Covid-19 cases in 34 provinces in Indonesia with the time period 1 July - 30 September 2021. Based on the copula-based RLTGKT model, the  $R^2$  value is 90.92%, KTG is 180.762 and MAPE is 63.10%. By using Clayton Copula, the GWTCLR Model can capture the different spatial and temporal influences of weather variable factors on the spread Covid-19 in Indonesia.

**Keywords:** Logistic Regression, Binomial Distribution, Spatial, Geographically weighted Regression, Temporal, Time Dependency, Clayton Copula, Gumbel Copula, Frank Copula.