

ABSTRAK

World Happiness Report merupakan laporan yang menyajikan hasil pengolahan survei internasional dalam bentuk indeks kebahagiaan negara-negara di dunia. Sejak tahun 2014, Badan Pusat Statistik (BPS) juga melakukan survei untuk membangun indeks kebahagiaan provinsi-provinsi di Indonesia. Survei tersebut dilakukan tiga tahun sekali, untuk menilai dimensi kepuasan hidup, afeksi, dan makna hidup. Berdasarkan data indikator kebahagiaan dan kesejahteraan sosial dari BPS untuk masing-masing provinsi di Indonesia serta pendapat para ahli, akan dilakukan pengelompokan provinsi menggunakan dua metode *clustering*, K-Means dan K-Medoids. Metode K-Medoids memberikan hasil yang lebih baik karena menghasilkan Variance Ratio Criterion yang cenderung lebih baik, *cluster* yang lebih terpisah, dan lebih dapat diinterpretasi berdasarkan nilai variabel-variabelnya. Hal ini menunjukkan keunggulan metode tersebut apabila terdapat banyak *outlier* pada data, sehingga dipilih untuk pembahasan selanjutnya. Dari metode K-Medoids, diperoleh 2 *cluster* kebahagiaan dan 3 *cluster* kesejahteraan. Provinsi yang bahagia cenderung berkumpul di daerah tengah, sedangkan provinsi yang sejahtera lebih tersebar. Karakteristik *cluster* diidentifikasi melalui *feature selection* dengan metode Ordinal Logistic Regression with Elastic Net Penalty dan Random Forest Classifier. Kedua model memiliki performa yang baik dari segi pemilihan variabel, di mana hanya tersisa kurang dari 40% variabel untuk data kebahagiaan dan kurang dari 50% untuk variabel kesejahteraan. Selain itu, kedua model juga memberikan akurasi yang baik dengan skor di atas 80%. Keunggulan OLR with Elastic Net Penalty dapat dilihat dari segi interpretabilitas karena mampu menunjukkan dampak positif dan negatif dari setiap variabel terhadap kebahagiaan dan kesejahteraan melalui koefisiennya, tetapi penalti Elastic Net mempersulit uji χ^2 untuk menghitung signifikansi variabel. Di lain sisi, RF mampu menunjukkan tingkat signifikansi masing-masing variabel berdasarkan Gini Importance. Hasil yang beririsan pada kedua metode menunjukkan bahwa dimensi makna hidup adalah yang terpenting untuk meningkatkan kebahagiaan. Sedangkan dimensi-dimensi yang dapat dipertimbangkan untuk pembangunan selanjutnya adalah dimensi teknologi, kesehatan, perumahan, dan kependudukan.

Kata Kunci: Tingkat Kebahagiaan, Kesejahteraan Sosial, Provinsi, *Clustering*, *Feature Selection*

ABSTRACT

The World Happiness Report provides a happiness index for each country in the world, which was built based on the results of their international survey. According to that, the Indonesian Central Bureau of Statistics (Badan Pusat Statistik Indonesia) also conducted a survey to determine each province's happiness. The survey was conducted once every three years, the first one being in 2014. The object of observation was life satisfaction, affection, and eudaimonia (life meaning). First, this research will focus on performing K-Means and K-Medoids clustering to group the provinces in Indonesia based on its happiness and social welfare. K-Medoids have proven to be advantageous when there are outliers within the dataset, offering better Variance Ratio Criterion, more separated clusters, and better interpretability in terms of variables. Thus, K-Medoids's results were chosen for the next part. 2 happiness clusters and 3 social welfare clusters were obtained, with the happier provinces tend group in the middle, while provinces with high social welfare were more distributed. To identify the characteristics of each cluster, two feature selection methods: Ordinal Logistic Regression (OLR) with Elastic Net Penalty and Random Forest Classifier (RF) will be used. Both models performed well on reducing the number of variables, leaving only 40% of the happiness variables and 50% of the social welfare variables. OLR with Elastic Net Penalty provides more interpretability, considering its capability of explaining the impact of each variable based on their coefficients. However, the penalty complicates the usual χ^2 significance test. RF on the other hand, is capable of showing the importance of each variable using Gini Importance. Based on the intersection of these methods, eudaimonia is the most important dimension that can improve happiness, while technology, health, property, and civic development are the dimensions that can be considered to improve social welfare.

Keywords: Happiness Rate, Social Welfare, Indonesia's Province, Clustering, Feature Selection