



PENGGUNAAN TEKNIK DATA MINING

DALAM ANALISIS IZIN TINGGAL KEIMIGRASIAN

Dr. Priati, S.Kom., M.Kom.
Besse Hartati, S.S., M.M.
Isidorus Anung Prabadhi, S.Sos., M.M.
Nurul Maharani Piranti, A.Md., S.T., M.Si.
Rasona Sunara Akbar, S.Pd., M.M
Galuh Boy Hertantyo, S.T., M.T.
Masdar Bakhtiar, S.H., M.H.
Atsil Syah Gibran
Tiaka Okta Damayanti

PENGGUNAAN TEKNIK DATA MINING

DALAM ANALISIS IZIN TINGGAL KEIMIGRASIAN

Dr. Priati, S.Kom., M.Kom.
Besse Hartati, S.S., M.M.
Isidorus Anung Prabadhi, S.Sos., M.M.
Nurul Maharani Piranti, A.Md., S.T., M.Si.
Rasona Sunara Akbar, S.Pd., M.M
Galuh Boy Hertantyo, S.T., M.T.
Masdar Bakhtiar, S.H., M.H.
Atsil Syah Gibran
Tiaka Okta Damayanti

**PENGGUNAAN TEKNIK DATA MINING
DALAM ANALISIS IZIN TINGGAL KEIMIGRASIAN**

Penulis:

Dr. Priati, S.Kom., M.Kom.
Besse Hartati, S.S., M.M.
Isidorus Anung Prabadhi, S.Sos., M.M.
Nurul Maharani Piranti, A.Md., S.T., M.Si.
Rasona Sunara Akbar, S.Pd., M.M
Galuh Boy Hertantyo, S.T., M.T.
Masdar Bakhtiar, S.H., M.H.
Atsil Syah Gibran
Tiaka Okta Damayanti

Desain Cover:

Septian Maulana

Sumber Ilustrasi:

www.freepik.com

Tata Letak:

Handarini Rohana

Editor:

Dr. Elan Jaelani, S.H., M.H

ISBN:

978-623-500-127-2

Cetakan Pertama:

April, 2024

Hak Cipta Dilindungi Oleh Undang-Undang

by Penerbit Widina Media Utama

Dilarang keras menerjemahkan, memfotokopi, atau memperbanyak sebagian atau seluruh isi buku ini tanpa izin tertulis dari Penerbit.

PENERBIT:

WIDINA MEDIA UTAMA

Komplek Puri Melia Asri Blok C3 No. 17 Desa Bojong Emas
Kec. Solokan Jeruk Kabupaten Bandung, Provinsi Jawa Barat

Anggota IKAPI No. 360/JBA/2020

Website: www.penerbitwidina.com

Instagram: @penerbitwidina

Telepon (022) 87355370

PRAKATA PENULIS

Assalamualaikum Warahmatullahi Wabarakatuh

Alhamdulillahirabbil'alamin, puji dan syukur kami panjatkan kehadirat Allah SWT atas berkat dan rahmat-Nya, buku dengan judul "*Penggunaan Teknik Data Mining Dalam Analisis Izin Tinggal Keimigrasian*" dapat kami selesaikan secara paripurna.

Buku ini lebih banyak membahas tentang pemanfaatan teknologi informasi yaitu *data mining* atau penambangan data untuk menemukan informasi penting pada data izin tinggal keimigrasian. Dewasa ini, teknik tersebut banyak digunakan sebagai support atau pendukung dalam pengambilan keputusan sebuah organisasi. Secara khusus teknik *data mining* yang diaplikasikan dalam buku ini adalah teknik pengklasteran atau *clustering*.

Sebagaimana teknologi lainnya, *data mining* bersifat sangat dinamis dan akan terus berkembang seiring dengan perkembangan teknologi komputasi, yang jika digambarkan dalam grafik selalu bersifat eksponensial. Artinya, grafik perkembangan teknologi terutama dalam bidang teknologi informasi dan komputer akan terus meningkat setiap waktu.

Menyadari hal tersebut, penulis berharap perbaikan-perbaikan dapat dilakukan dikemudian hari untuk menyesuaikan perkembangan teknologinya. Oleh karena itu penulis sangat antusias terhadap kritik dan saran dari pembaca.

Melalui buku ini kami haturkan terima kasih yang tak terhingga terhadap semua pihak yang telah mendukung dalam penyusunan hingga penyelesaiannya.

Jakarta, 27 Oktober 2023

Tim Penulis

DAFTAR ISI

PRAKATA PENULIS	iii
DAFTAR ISI	v
BAB 1 PELAKSANAAN KEIMIGRASIAN DI INDONESIA	1
BAB 2 RELEVANSI DATA MINING DALAM ANALISIS IZIN TINGGAL KEIMIGRASIAN	5
A. Pengertian <i>Data Mining</i>	5
B. Teknik Utama Dalam <i>Data Mining</i>	9
C. Relevansi <i>Data Mining</i> dan Izin Tinggal	15
BAB 3 METODE CRISP-DM DALAM ANALISIS IZIN TINGGAL KEIMIGRASIAN	19
BAB 4 IMPLEMENTASI DATA MINING DALAM PROSES ANALISIS IZIN TINGGAL DI KANTOR IMIGRASI	29
A. Fase Pemahaman Bisnis	29
B. Fase Pemahaman Data	30
C. Fase Persiapan Data	32
D. <i>Modeling</i>	49
E. <i>Evaluation</i>	52
F. <i>Deployment</i>	52
DAFTAR PUSTAKA	54



BAB
1

**PELAKSANAAN KEIMIGRASIAN
DI INDONESIA**

Sejak zaman dahulu, perpindahan manusia dari lokasi satu ke lokasi lain, yang dikenal dengan istilah migrasi, telah menjadi fenomena umum. Adanya faktor yang mendorong dan faktor yang menarik merupakan alasan di balik terjadinya fenomena tersebut. Alasan utama yang sering menjadi dorongan adalah untuk mencari pekerjaan (dikenal dengan teori faktor dorongan dan faktor tarikan) (Lee, 1966). Faktor-faktor yang mempengaruhi terjadinya migrasi internasional meliputi faktor ekonomi, faktor sosial budaya, dan faktor stabilitas politik. Indonesia, sebagai negara berkembang dengan luas wilayah dan lebih dari 17.500 pulau, menjadi negara kepulauan terbesar di dunia dan menghadapi berbagai tantangan terkait populasi, termasuk pertumbuhan populasi yang tinggi dan ketidakmerataan distribusi penduduk. Pergerakan penduduk baik di dalam negeri maupun ke luar negeri menjadi suatu hal yang umum terjadi di Indonesia. Selain itu, Indonesia juga menarik perhatian sebagai tujuan migrasi penduduk dari negara lain, hal ini terjadi karena keindahan alam dan kekayaan alam yang melimpah di Indonesia.



RELEVANSI DATA MINING DALAM ANALISIS IZIN TINGGAL KEIMIGRASIAN

A. PENGERTIAN *DATA MINING*

Data mining atau penambangan data dapat diartikan sebagai suatu proses untuk mengidentifikasi hubungan pola dan tren baru yang memiliki makna dengan memanfaatkan teknik pengenalan pola seperti teknik statistik dan matematika (Sato et al., 2019)(Gola et al., 2018). Istilah *data mining* digunakan untuk merujuk pada penemuan pengetahuan dalam basis data. *Data mining* merupakan proses yang melibatkan penerapan metode analisis (Setiawidayat & Yuniar Rahman, 2018), konsep matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengekstrak serta mengidentifikasi informasi yang berharga dan pengetahuan terperinci dari berbagai sumber basis data besar (H. Wu et al., 2018). Buku (Larose & Larose, n.d.) tentang konsep dasar dan metode dalam penambangan data, membantu pembaca memahami bagaimana mengaplikasikan alat dan teknik data mining dalam analisis data yang beragam.

Teknik *data mining* merupakan pendekatan yang kuat untuk menggali informasi berharga dari data besar dan kompleks. Ini melibatkan penggunaan algoritma komputasional untuk



METODE CRISP-DM DALAM ANALISIS IZIN TINGGAL KEIMIGRASIAN

Izin tinggal keimigrasi merupakan wewenang yang diberikan kepada orang asing untuk tinggal di wilayah Indonesia dalam jangka waktu tertentu, dengan aktivitas yang sejalan dengan jenis visa yang dimiliki oleh orang asing. Pemberian Izin Tinggal Keimigrasian merupakan bagian dari layanan publik yang diberikan oleh sektor imigrasi kepada masyarakat, terutama kepada orang asing yang hendak memasuki, tinggal, dan beraktivitas di wilayah Indonesia dalam batasan waktu tertentu.

Kategori-kategori Izin Tinggal Keimigrasian yang sesuai dengan yang tertera dalam Undang-Undang Nomor 6 Tahun 2011 (*Undang-Undang Republik Indonesia Nomor 6 Tahun 2011, n.d.*) tentang keimigrasian mencakup Izin Tinggal Diplomatik, Izin Tinggal Dinas, Izin Tinggal Kunjungan, Izin Tinggal Terbatas, dan Izin Tinggal Tetap.

Pengalihan status izin tinggal imigrasi digunakan sebagai alat untuk mengubah status izin tinggal imigrasi seseorang dari satu jenis izin tinggal ke jenis lainnya, seperti dari izin tinggal kunjungan menjadi izin tinggal terbatas, atau dari izin tinggal terbatas menjadi izin tinggal tetap.



IMPLEMENTASI DATA MINING DALAM PROSES ANALISIS IZIN TINGGAL DI KANTOR IMIGRASI

A. FASE PEMAHAMAN BISNIS

Kantor Imigrasi Kelas II Non TPI Kediri memiliki wilayah kerja yang cukup luas sehingga diperlukan adanya menganalisis data untuk dijadikan pola pengetahuan, untuk melakukan analisis terkait karakteristik pemohon Izin Tinggal Keimigrasian yang nantinya dapat digunakan dalam membantu perumusan kebijakan dan pengambilan keputusan, Diharapkan proses klastering dapat menemukan data yang detail dan terstruktur sehingga dapat dijadikan komponen pendukung dalam penentuan kebijakan selanjutnya terkait pemberian izin tinggal dan dalam rangka monitoring keberadaan orang asing dan dapat dibuat peta kerawanan orang asing. Dengan demikian pejabat dan pegawai dengan mudah dan cepat dalam mengetahui bagaimana karakteristik pemohon izin tinggal keimigrasian. Langkah awal yang digunakan ialah dengan mencari data penerbitan Izin Tinggal Keimigrasian Kantor Imigrasi Kelas II Non TPI Kediri yang tersedia pada Sistem Informasi Manajemen Keimigrasian (SIMKIM). Data yang digunakan pada penelitian ini ialah Data Penerbitan Izin

DAFTAR PUSTAKA

- Ahmad, A., & Khan, S. S. (2019). Survey of State-of-the-Art Mixed Data Clustering Algorithms. *IEEE Access*, 7, 31883–31902. <https://doi.org/10.1109/ACCESS.2019.2903568>
- Cagnone, S., & Viroli, C. (2014). A factor mixture model for analyzing heterogeneity and cognitive structure of dementia. *AStA Advances in Statistical Analysis*, 98(1), 1–20. <https://doi.org/10.1007/s10182-012-0206-5>
- Chapman, Pete, Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (1999). *CRISP-DM 1.0 Step-by-step data mining guide*. DaimlerChrysler.
- Chapman, Peter. (1999). The CRISP-DM User Guide. *The CRISP-DM User Guide*, 14.
- Chiu, T., Fang, D. P., Chen, J., Wang, Y., & Jeris, C. (2001). A robust and scalable clustering algorithm for mixed type attributes in large database environment. *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 263–268. <https://doi.org/10.1145/502512.502549>
- Dutt, A., Ismail, M. A., & Herawan, T. (2017). A Systematic Review on Educational Data Mining. In *IEEE Access* (Vol. 5, pp. 15991–16005). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ACCESS.2017.2654247>
- Fuchs, R., Pommeret, D., & Viroli, C. (2020). *Mixed data Deep Gaussian Mixture Model: A clustering model for mixed datasets*. <https://doi.org/10.1007/s11634-021-00466-3>

- Gao, Y., Miao, X., Chen, G., Zheng, B., Cai, D., & Cui, H. (2017). On efficiently finding reverse k-nearest neighbors over uncertain graphs. *VLDB Journal*, 26(4), 467–492. <https://doi.org/10.1007/s00778-017-0460-y>
- Gola, J., Britz, D., Staudt, T., Winter, M., Schneider, A. S., Ludovici, M., & Mücklich, F. (2018). Advanced microstructure classification by data mining methods. *Computational Materials Science*, 148, 324–335. <https://doi.org/10.1016/j.commatsci.2018.03.004>
- Greggio, N., Bernardino, A., Laschi, C., Dario, P., & Santos-Victor, J. (2012). Fast estimation of Gaussian mixture models for image segmentation. *Machine Vision and Applications*, 23(4), 773–789. <https://doi.org/10.1007/s00138-011-0320-5>
- Huang, Z. (1998). Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values. *Data Mining and Knowledge Discovery* 2, 1(3), 284–290. <https://doi.org/10.3923/ajbmb.2011.284.290>
- Kohonen, T. (1990). The Self-Organizing Map. *Proceedings of the IEEE*, 78(9), 1464–1480. <https://doi.org/10.1109/5.58325>
- Larose, D. T., & Larose, C. D. (n.d.). *Discovering knowledge in data: an introduction to data mining*.
- Lee, E. S. (1966). A Theory of Migration Author (s): Everett S . Lee
Published by: Springer on behalf of the Population Association of America Stable URL:
<http://www.jstor.org/stable/2060063> Accessed: 05-06-2016 17: 53 UTC Your use of the JSTOR archive indicates your.
Demography, 3(1), 47–57.
- Lia Hananto, A., Assiroj, P., Priyatna, B., Nurhayati, Fauzi, A., Yuniar Rahman, A., & Shofiah Hilabi, S. (2021). Analysis of Drug Data

- Mining with Clustering Technique Using K-Means Algorithm. *Journal of Physics: Conference Series*, 1908(1).
<https://doi.org/10.1088/1742-6596/1908/1/012024>
- Liu, H., Shao, M., Li, S., & Fu, Y. (2016). Infinite ensemble for image clustering. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-August-2016*, 1745–1754.
<https://doi.org/10.1145/2939672.2939813>
- Liu, X., Wan, C., & Chen, L. (2011). Returning clustered results for keyword search on XML documents. *IEEE Transactions on Knowledge and Data Engineering*, 23(12), 1811–1825.
<https://doi.org/10.1109/TKDE.2011.183>
- Lockwood, M. (2010). Good governance for terrestrial protected areas: A framework, principles and performance outcomes. *Journal of Environmental Management*, 91(3), 754–766.
<https://doi.org/10.1016/j.jenvman.2009.10.005>
- Marbán, O., Segovia, J., Menasalvas, E., & Fernández-Baizán, C. (2009). Toward data mining engineering: A software engineering approach. *Information Systems*, 34(1), 87–107.
<https://doi.org/10.1016/j.is.2008.04.003>
- Mariscal, G., Marbán, Ó., & Fernández, C. (2010). A survey of data mining and knowledge discovery process models and methodologies. *Knowledge Engineering Review*, 25(2), 137–166. <https://doi.org/10.1017/S0269888910000032>
- Martinez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernandez-Orallo, J., Kull, M., Lachiche, N., Ramirez-Quintana, M. J., & Flach, P. (2021). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. *IEEE*

- Transactions on Knowledge and Data Engineering*, 33(8), 3048–3061. <https://doi.org/10.1109/TKDE.2019.2962680>
- Miner Tulsa, G., Nisbet, R. A., Thompson Woodward, J., Richard Foley Raleigh, U., Balakrishnan, K., & Boston Heidelberg London New York Oxford Paris San Diego San Francisco Singapore Sydney Tokyo, A. (n.d.). *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications*. <http://elsevier.com>
- Moustaki, I. (2000). *Generalized latent trait models*. 65(3), 391–411.
- Moustaki, I. (2003). A general class of latent variable models for ordinal manifest variables with covariate effects on the manifest and latent variables. *British Journal of Mathematical and Statistical Psychology*, 56(2), 337–357. <https://doi.org/10.1348/000711003770480075>
- Philip, G., & Ottaway, B. S. (1983). Mixed Data Cluster Analysis: an Illustration Using Cypriot Hooked-Tang Weapons. *Archaeometry*, 25(2), 119–133. <https://doi.org/10.1111/j.1475-4754.1983.tb00671.x>
- Sato, Y., Izui, K., Yamada, T., & Nishiwaki, S. (2019). Data mining based on clustering and association rule analysis for knowledge discovery in multiobjective topology optimization. *Expert Systems with Applications*, 119, 247–261. <https://doi.org/10.1016/j.eswa.2018.10.047>
- Setiawidayat, S., & Yuniar Rahman, A. (2018). *New method for obtaining Peak Value R and the duration of each cycle of Electrocardiogram*. IEEE.
- Sumitra, S. (n.d.). *Mixture Models and EM Algorithm*.
- Taylor, P., Fraley, C., & Raftery, A. E. (2012). *Journal of the American Statistical Association and Density Estimation Model-Based*

Clustering , Discriminant Analysis , and Density Estimation.
May, 37–41.

- Undang-Undang Republik Indonesia Nomor 6 Tahun 2011. (n.d.).
- Wang, Y., Lin, X., Wu, L., Zhang, W., & Zhang, Q. (2014). Exploiting correlation consensus: Towards subspace clustering for multi-modal data. *MM 2014 - Proceedings of the 2014 ACM Conference on Multimedia*, 981–984.
<https://doi.org/10.1145/2647868.2654999>
- Weber, C. M., Ray, D., Valverde, A. A., Clark, J. A., & Sharma, K. S. (2021). *Gaussian mixture model clustering algorithms for the analysis of high-precision mass measurements.*
<https://doi.org/10.1016/j.nima.2021.166299>
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data Mining: Practical Machine Learning Tools and Techniques. In *Data Mining: Practical Machine Learning Tools and Techniques.*
<https://doi.org/10.1016/c2009-0-19715-5>
- Wu, H., Yang, S., Huang, Z., He, J., & Wang, X. (2018). Type 2 diabetes mellitus prediction model based on data mining. *Informatics in Medicine Unlocked*, 10, 100–107.
<https://doi.org/10.1016/j.imu.2017.12.006>
- Wu, X., Kumar, V., Ross, Q. J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, P. S., Zhou, Z. H., Steinbach, M., Hand, D. J., & Steinberg, D. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems.*
<https://doi.org/10.1007/s10115-007-0114-2>
- Han, J., Kamber, M., & Pei, J. (2011). *Data mining: concepts and techniques* (3rd ed.). Elsevier.

- Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM Computing Surveys (CSUR)*, 31(3), 264-323.
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651-666.
- Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, 50(2), 159-179.
- Xu, R., & Tian, Y. (2015). A comprehensive survey of clustering algorithms. *Annals of Data Science*, 2(2), 165-193.
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques*. Morgan Kaufmann.
- Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.
- Tan, P. N., Steinbach, M., & Kumar, V. (2018). *Introduction to Data Mining*. Pearson.
- Witten, I. H., Frank, E., & Hall, M. A. (2016). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann

PENGUNAAN TEKNIK DATA MINING

DALAM ANALISIS IZIN TINGGAL KEIMIGRASIAN

Sejak zaman dahulu, perpindahan manusia dari lokasi satu ke lokasi lain, yang dikenal dengan istilah migrasi, telah menjadi fenomena umum. Adanya faktor yang mendorong dan faktor yang menarik merupakan alasan di balik terjadinya fenomena tersebut. Alasan utama yang sering menjadi dorongan adalah untuk mencari pekerjaan (dikenal dengan teori faktor dorongan dan faktor tarikan). Faktor-faktor yang mempengaruhi terjadinya migrasi internasional meliputi faktor ekonomi, faktor sosial budaya, dan faktor stabilitas politik. Indonesia, sebagai negara berkembang dengan luas wilayah dan lebih dari 17.500 pulau, menjadi negara kepulauan terbesar di dunia dan menghadapi berbagai tantangan terkait populasi, termasuk pertumbuhan populasi yang tinggi dan ketidakmerataan distribusi penduduk. Pergerakan penduduk baik di dalam negeri maupun ke luar negeri menjadi suatu hal yang umum terjadi di Indonesia. Selain itu, Indonesia juga menarik perhatian sebagai tujuan migrasi penduduk dari negara lain, hal ini terjadi karena keindahan alam dan kekayaan alam yang melimpah di Indonesia.

Oleh karena itu, buku ini bertujuan untuk menginvestigasi potensi pemanfaatan teknik Data Mining dalam analisis data izin tinggal keimigrasian kantor imigrasi kelas II Non TPI Kediri. Dengan menggali informasi yang tersembunyi dalam data, diharapkan buku ini dapat memberikan kontribusi dalam meningkatkan pengelolaan data izin tinggal, mendukung pengambilan keputusan yang lebih baik, serta mendeteksi potensi risiko dan ancaman keamanan nasional. Dengan demikian, buku ini memiliki relevansi yang tinggi dalam menghadapi tantangan pengelolaan data perlintasan WNA dan keamanan nasional di era digital ini. Selamat membaca!