

EVALUASI PERFORMA *ENSEMBLE* SVM BERBASIS *FUZZY K-PROTOTYPES* PADA DATA PENGANGGURAN TERBUKA DI PROVINSI JAWA BARAT

SETYO WAHYUDI



**PROGRAM STUDI MAGISTER STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



PERNYATAAN MENGENAI TESIS DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa tesis dengan judul “Evaluasi Performa *Ensemble SVM* Berbasis *Fuzzy K-Prototypes* pada Data Pengangguran Terbuka di Provinsi Jawa Barat” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir tesis ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Agustus 2026

Setyo Wahyudi
G1501222046

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



RINGKASAN

SETYO WAHYUDI. Evaluasi Performa *Ensemble SVM* Berbasis *Fuzzy K-Prototype* pada Data Pengangguran Terbuka di Provinsi Jawa Barat. Dibimbing oleh **INDAHWATI** dan **I MADE SUMERTAJAYA**.

Penelitian ini dilatarbelakangi oleh permasalahan klasifikasi pada data yang tidak seimbang (*class imbalance*), di mana jumlah observasi pada kelas minoritas jauh lebih sedikit dibandingkan kelas mayoritas. Pada kondisi tersebut, model klasifikasi konvensional cenderung menghasilkan akurasi yang tinggi, tetapi memiliki kemampuan yang rendah dalam mendeteksi kelas minoritas. Permasalahan ini menjadi penting pada kasus pengangguran terbuka di Provinsi Jawa Barat karena proporsi pengangguran hanya sekitar 6,5% dari seluruh angkatan kerja, sehingga individu yang menganggur lebih sulit dikenali oleh model klasifikasi.

Support Vector Machine (SVM) merupakan salah satu algoritma klasifikasi yang memiliki kemampuan generalisasi yang baik, namun performanya menurun ketika diterapkan pada data yang sangat tidak seimbang. Berbagai penelitian telah menunjukkan bahwa penerapan teknik *oversampling* seperti SMOTE mampu meningkatkan representasi kelas minoritas. Akan tetapi, SMOTE hanya sesuai untuk data numerik, sedangkan data ketenagakerjaan terdiri atas kombinasi peubah numerik, nominal, dan ordinal sehingga memerlukan pendekatan yang lebih sesuai, yaitu SMOTE-NC.

Efektivitas SMOTE-NC yang digunakan untuk penyeimbangan data sangat dipengaruhi oleh struktur data yang digunakan dalam pembentukan data sintetis. Oleh karena itu, penelitian ini mengusulkan proses penggerombolan sebelum *oversampling* menggunakan metode *Fuzzy K-Prototype* (FKP). Berbeda dengan *K-Means* yang hanya dapat menangani data numerik dan menggunakan pendekatan *hard clustering*, FKP mampu mengelompokkan data campuran dengan konsep *fuzzy membership*, sehingga setiap observasi dapat memiliki derajat keanggotaan pada lebih dari satu gerombol dan menghasilkan kelompok yang lebih homogen.

Penelitian ini mengembangkan kerangka kerja *clustering-based SVM ensemble* dengan mengganti metode *K-Means* menjadi *Fuzzy K-Prototype* dan mengganti SMOTE menjadi SMOTE-NC. Proses yang diawali dengan menggunakan metode penggerombolan kemudian dilihat apakah kelas dalam kondisi tidak seimbang, jika kelas dalam kondisi tidak seimbang maka dilakukan penyeimbangan kelas terlebih dahulu pada setiap gerombol yang terbentuk menggunakan SMOTE-NC. Model *ensemble SVM* akan dilakukan setelah kondisi kelas data sudah seimbang pada masing-masing gerombol. Hasil prediksi akhir diperoleh melalui proses *ensemble* yang mempertimbangkan bobot performa setiap model serta derajat keanggotaan setiap observasi terhadap seluruh gerombol.

Data yang digunakan berasal dari Survei Angkatan Kerja Nasional (Sakernas) Agustus 2024 Provinsi Jawa Barat dengan jumlah observasi sebanyak 30.755 individu pada kategori usia kerja. Peubah yang digunakan terdiri atas kombinasi peubah numerik, nominal, dan ordinal yang meliputi usia, jenis kelamin, tingkat pendidikan, status perkawinan, status dalam rumah tangga, keikutsertaan Program Kartu Prakerja, riwayat pelatihan, dan lama masa setelah lulus pendidikan. Evaluasi model dilakukan menggunakan sensitivitas sebagai nilai evaluasi utama, dan juga



nilai *balanced accuracy*, spesifisitas, dan akurasi. Hasil analisis *Fuzzy K-Prototype* menunjukkan bahwa jumlah gerombol optimal adalah dua berdasarkan nilai silhouette score tertinggi.

Gerombol pertama didominasi oleh laki-laki, kepala keluarga, berusia lebih tua, dan memiliki tingkat pengangguran yang rendah. Sebaliknya, gerombol kedua didominasi oleh perempuan, bukan kepala keluarga, berusia lebih muda, serta memiliki proporsi pengangguran yang jauh lebih tinggi sehingga lebih rentan terhadap risiko pengangguran. Evaluasi terhadap model *ensemble SVM* menunjukkan bahwa tanpa penanganan *class imbalance*, model menghasilkan akurasi yang tinggi namun sensitivitas sangat rendah sehingga gagal mendeteksi sebagian besar individu yang menganggur. Setelah diterapkan SMOTE-NC, kemampuan model dalam mengenali kelas minoritas meningkat secara signifikan. Hasil ini menunjukkan bahwa penyeimbangan data merupakan tahapan yang sangat penting sebelum proses klasifikasi dilakukan pada data ketenagakerjaan yang tidak seimbang.

Penerapan metode *ensemble* berbasis *Fuzzy K-Prototype* dan SMOTE-NC menghasilkan peningkatan performa klasifikasi dibandingkan model SVM. Penggabungan antara proses penggerombolan, *oversampling*, dan *ensemble learning* mampu menghasilkan model yang lebih stabil, meningkatkan sensitivitas dalam mendeteksi pengangguran, serta tetap mempertahankan nilai *balanced accuracy* yang baik. Selain itu, evaluasi terhadap jumlah gerombol menunjukkan bahwa kualitas penggerombolan berdasarkan *silhouette score* dan metode *elbow* berpengaruh terhadap performa klasifikasi, khususnya sensitivitas.

Secara keseluruhan, penelitian ini menunjukkan bahwa pendekatan *ensemble SVM* berbasis *Fuzzy K-Prototype* merupakan metode yang efektif untuk menangani data dengan peubah campuran yang tidak seimbang. Kombinasi metode tersebut mampu menghasilkan data sintetis yang lebih representatif, meningkatkan kemampuan model dalam mengenali kelas minoritas, dan memberikan performa klasifikasi yang lebih baik dibandingkan pendekatan model SVM. Temuan penelitian ini diharapkan dapat menjadi alternatif metode klasifikasi pada data bertipe campuran serta mendukung penyusunan kebijakan ketenagakerjaan yang lebih tepat sasaran.

Kata kunci: *ensemble learning, fuzzy k-prototypes, imbalanced data, pengangguran terbuka, SMOTE-NC, support vector machine.*



SUMMARY

SETYO WAHYUDI. Performance Evaluation of a Fuzzy K-Prototype-Based Ensemble SVM on Open Unemployment Data in West Java Province. Supervised by **INDAHWATI** and **I MADE SUMERTAJAYA**.

This study was motivated by the classification challenges posed by imbalanced data, where the minority class contains substantially fewer observations than the majority class. Under such conditions, conventional classification models often achieve high overall accuracy while performing poorly in identifying minority-class instances. This issue is particularly important in the case of open unemployment in West Java Province, where unemployed individuals represent only about 6.5% of the labor force, making them difficult to detect accurately.

Support Vector Machine (SVM) is a widely used classification algorithm because of its strong generalization capability. However, its performance deteriorates when applied to highly imbalanced datasets. Previous studies have shown that oversampling techniques such as SMOTE can improve minority-class representation. Nevertheless, conventional SMOTE is designed only for numerical data, whereas labor force datasets consist of mixed numerical, nominal, and ordinal variables, making SMOTE-NC a more appropriate alternative.

The effectiveness of SMOTE-NC largely depends on the underlying data structure used to generate synthetic samples. Therefore, this study proposes a clustering stage prior to oversampling using the Fuzzy K-Prototype (FKP) algorithm. Unlike K-Means, which is limited to numerical data and employs hard clustering, FKP is specifically designed for mixed-type data and utilizes fuzzy memberships, allowing each observation to belong to multiple clusters with different membership degrees and producing more homogeneous clusters.

This research extends the clustering-based SVM ensemble framework by replacing K-Means with Fuzzy K-Prototype and conventional SMOTE with SMOTE-NC. After clustering, SMOTE-NC is applied independently within each cluster to balance the class distribution, followed by the construction of an SVM classifier for each cluster. The final prediction is generated through an ensemble mechanism that integrates classifier confidence weights and the fuzzy membership degree of each observation.

The study employed data from the August 2024 National Labor Force Survey (Sakernas) in West Java, comprising 30,755 working-age individuals. The dataset included numerical, nominal, and ordinal variables such as age, gender, education level, marital status, household status, participation in the Pre-Employment Card Program, training history, and years since graduation. Model performance was primarily evaluated using sensitivity, complemented by balanced accuracy, specificity, and overall accuracy. The Fuzzy K-Prototype clustering results indicated that two clusters provided the optimal solution according to the highest silhouette score.

The first cluster was predominantly composed of older male household heads with a very low unemployment rate. In contrast, the second cluster mainly consisted of younger females who were not household heads and exhibited a substantially higher unemployment rate, indicating greater vulnerability to unemployment. The evaluation of the conventional SVM model revealed that, without addressing class



- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

imbalance, the model achieved high overall accuracy but extremely low sensitivity, failing to identify most unemployed individuals. After applying SMOTE-NC, the model's ability to recognize the minority class improved considerably. These findings demonstrate that class balancing is an essential preprocessing step for labor force datasets with severe class imbalance.

The proposed ensemble model based on Fuzzy K-Prototype and SMOTE-NC outperformed the SVM model. The integration of clustering, oversampling, and ensemble learning produced a more robust classifier, substantially improved sensitivity in detecting unemployed individuals, and maintained satisfactory balanced accuracy. Furthermore, the evaluation of different numbers of clusters indicated that clustering quality, measured by the silhouette score dan elbow method, had a meaningful impact on classification performance, particularly sensitivity.

Overall, this study demonstrates that the Fuzzy K-Prototype-based Ensemble-SVM framework is an effective approach for handling imbalanced mixed-type data. The proposed method generates more representative synthetic samples, enhances the detection of minority-class observations, and achieves superior classification performance compared with conventional approaches. These findings contribute to the development of machine learning methods for mixed-type socioeconomic data and provide valuable support for evidence-based employment policy and unemployment mitigation strategies.

Keywords: *ensemble learning, fuzzy k-prototypes, imbalanced data, open unemployment, SMOTE-NC, support vector machine*



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2026
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.

EVALUASI PERFORMA *ENSEMBLE* SVM BERBASIS *FUZZY K-PROTOTYPES* PADA DATA PENGANGGURAN TERBUKA DI PROVINSI JAWA BARAT

SETYO WAHYUDI

Tesis
sebagai salah satu syarat untuk memperoleh gelar
Magister Sains pada
Program Studi Statistika dan Sains Data

**PROGRAM STUDI MAGISTER STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



Judul Tesis : Evaluasi Performa *Ensemble SVM* Berbasis *Fuzzy K-Prototypes*
pada Data Pengangguran Terbuka di Provinsi Jawa Barat

Nama : Setyo Wahyudi

NIM : G1501222046

Disetujui oleh

Pembimbing 1:

1. Dr. Ir. Indahwati, M.Si.

Pembimbing 2:

2. Prof. Dr. Ir. I Made Sumertajaya, M.Si.

Diketahui oleh

Ketua Program Studi:

Dr. Agus Mohamad Soleh, S.Si., M.T.

NIP 197503151999031004

Dekan Sekolah Sains Data, Matematika, dan Informatika

Prof. Dr. Ir. Agus Buono, M.Si., M.Kom.

NIP 196607021993021001

Tanggal Ujian: 08 Agustus 2026

Tanggal Lulus:



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga tesis dengan judul "Evaluasi Performa *Ensemble SVM* Berbasis *Fuzzy K-Prototype* pada Data Pengangguran Terbuka di Provinsi Jawa Barat" berhasil diselesaikan. Tesis ini penulis susun sebagai salah satu syarat untuk menyelesaikan studi Magister Sains pada Program Studi Statistika dan Sains Data.

Penulisan tesis ini dapat terselesaikan tidak terlepas dari dukungan, bimbingan, dan bantuan berbagai pihak sehingga penulis menyampaikan ucapan terima kasih kepada:

1. Dr. Ir. Indahwati, M.Si. dan Prof. Dr. Ir. I Made Sumertajaya, M.Si. selaku komisi pembimbing yang senantiasa bersedia meluangkan waktunya untuk memberikan bimbingan, masukan, dukungan, dan arahan kepada penulis;
2. Orangtua, istri dan kedua putri penulis yang senantiasa mendoakan, mendukung, dan memotivasi penulis;
3. Badan Pusat Statistik (BPS) yang telah bersedia menyediakan dan memberikan data hasil Survei Angkatan Kerja Nasional (Sakernas) bulan Agustus tahun 2024;
4. Seluruh dosen pengajar Departemen Statistika atas ilmu yang telah diberikan;
5. Seluruh Staf Tata Usaha, Departemen Statistika yang membantu dalam segala proses administrasi;
6. Rekan-rekan mahasiswa Pascasarjana Program Studi Statistika dan Sains Data angkatan 2022 atas dukungan dan kerjasama yang baik.

Semoga penelitian ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, Agustus 2026
Setyo Wahyudi

DAFTAR ISI

DAFTAR ISI	ii
DAFTAR TABEL	iv
DAFTAR GAMBAR	v
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Tujuan	6
1.4 Manfaat	6
1.5 Ruang Lingkup	6
II TINJAUAN PUSTAKA	7
2.1 Fuzzy K-Prototypes	7
2.2 Silhouette Score	10
2.3 <i>Synthetic Minority Over-sampling Technique for Nominal and Continuous</i> (SMOTE-NC)	10
2.4 <i>Support Vector Machine</i> (SVM)	12
2.5 <i>Ensemble SVM</i>	13
2.6 Evaluasi Model	14
III METODE PENELITIAN	16
3.1 Data	16
3.2 Tahapan Analisis Data	17
IV HASIL DAN PEMBAHASAN	19
4.1 Eksplorasi Data	19
4.2 Analisis Gerombol <i>Fuzzy K-Prototypes</i> (FKP)	22
4.3 Penanganan Data Tidak Seimbang	25
4.4 Model <i>Support Vector Machine</i> (SVM)	26
4.4.1 Model SVM Tanpa Penyeimbangan Kelas Data	26
4.4.2 Model SVM Dengan Penyeimbangan Kelas Data	28
4.5 Model <i>Ensemble SVM</i>	29
4.5.1 Model <i>Ensemble SVM</i> Tanpa SMOTE-NC	29
4.5.2 Model <i>Ensemble SVM</i> dengan SMOTE-NC	32
4.6 Perbandingan Model SVM dan Model <i>Ensemble SVM</i>	35
4.7 Evaluasi Pengaruh Jumlah Gerombol Pada Model <i>Ensemble SVM</i>	37

V	SIMPULAN DAN SARAN	39
5.1	Simpulan	39
5.2	Saran	39
	DAFTAR PUSTAKA	40
	RIWAYAT HIDUP	43



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

DAFTAR TABEL

1	Komposisi kegiatan di pasar kerja Provinsi Jawa Barat 2024	5
2	Matriks konfusi	15
3	Data peubah yang digunakan	16
4	Nilai hyperparameter SVM	17
5	Demografi gerombol berdasarkan nilai derajat keanggotaan terbesar	24
6	Distribusi kelas klasifikasi	25
7	Evaluasi model SVM pada data uji tanpa SMOTE-NC	27
8	Evaluasi model SVM pada data uji tanpa SMOTE-NC	29
9	Nilai evaluasi dan kontribusi bobot setiap model uji ensemble SVM tanpa SMOTE-NC pada masing-masing gerombol	31
10	Evaluasi model ensemble SVM	32
11	Nilai evaluasi dan kontribusi bobot setiap model uji ensemble SVM dengan SMOTE-NC pada masing-masing gerombol	34
12	Evaluasi model ensemble SVM	35
13	Perbandingan nilai evaluasi model	36

DAFTAR GAMBAR

1	Kategori penduduk menurut kegiatan pasar kerja di Indonesia	5
2	Konsep hyperplane	13
3	Tahapan dan alur analisis	18
4	Proporsi kategori angkatan kerja di Jawa Barat	19
5	Proporsi kategori angkatan kerja di Jawa Barat berdasarkan jenis kelamin	20
6	Proporsi kategori angkatan kerja di Jawa Barat berdasarkan tingkat pendidikan	20
7	Proporsi kategori angkatan kerja di Jawa Barat berdasarkan keikutsertaan program kartu prakerja	21
8	Proporsi kategori angkatan kerja di Jawa Barat berdasarkan lama masa setelah lulus pendidikan	22
9	Evaluasi kualitas hasil penggerombolan dengan <i>silhouette dan elbow</i>	22
10	Proporsi jumlah data hasil penggerombolan	23
11	Nilai evaluasi SVM pada data latih tanpa SMOTE-NC	27
12	Nilai evaluasi SVM pada data latih dengan SMOTE-NC	28
13	Nilai evaluasi model <i>ensemble</i> SVM tanpa SMOTE pada data gerombol 1	30
14	Nilai evaluasi model <i>ensemble</i> SVM tanpa SMOTE pada data gerombol 2	31
15	Nilai evaluasi model <i>ensemble</i> SVM pada data gerombol 1	33
16	Nilai evaluasi model <i>ensemble</i> SVM pada data gerombol 2	34
17	Evaluasi jumlah gerombol terhadap nilai sensitivitas pada metode <i>ensemble</i> SVM	37