



Deteksi Alergen pada Produk Pangan Menggunakan Algoritma *Support Vector Machines* (SVM)

Siska Narulita^{1*}, Sekarlangit Sekarlangit², Milka Putri Novianingrum³
¹⁻³Universitas Nasional Karangturi Semarang, Indonesia

Alamat: Jl. Raden Patah No. 182-192 Rejomulyo, Semarang Timur, Kota Semarang
Korespondensi penulis: siskanarulita84@gmail.com*

Abstract. Food allergies are medical conditions caused by particular immunological reactions brought on by exposure to certain foods. All age groups can experience food allergies, albeit the prevalence varies between children and adults, with children experiencing this condition more frequently than adults. Find food ingredients or substances that can trigger allergies, often known as allergens. This project attempts to determine whether or not the food includes allergies by applying the SVM data mining method to a public dataset of food goods and allergens that was acquired via Kaggle. High accuracy, effective memory use, and the ability to handle non-normally distributed data are some of the benefits of the SVM method. Data collection is the first step in the research process. Data pre-processing, which includes data transformation, handling missing values, and copy objects, comes next. Validation comes next. Split validation with 90% training data and 10% testing data, 10-fold cross validation, and split validation with an 80%–20% ratio were all compared in this study. The SVM method is applied after the dataset has passed validation, and the confusion matrix is used for the last evaluation step. SVM has an accuracy rate of 97.24% when using 10-fold cross validation, according to the accuracy value produced by the validation process comparison. Split validation yields an accuracy value of 97.50% when the ratio of training data to testing data is 90% to 10%. In contrast, an accuracy rate of 98.75% was achieved by using split validation with a ratio of 80% and 20%.

Keywords: Allergen, Data Mining, K-Fold Cross Validation, Split Validation, Support Vector Machines

Abstrak Alergi makanan merupakan gangguan kesehatan yang muncul diakibatkan oleh respon imun spesifik yang muncul dari paparan makanan. Alergi makanan ini dapat terjadi pada semua kelompok usia, dengan prevalensi yang berbeda antara anak dan dewasa, dimana prevalensi pada anak lebih besar dibandingkan dengan orang dewasa. Identifikasi alergen, yaitu bahan pangan atau senyawa yang dapat menyebabkan alergi. Menggunakan *dataset* publik produk pangan dan alergen yang diperoleh dari Kaggle, penelitian ini bertujuan untuk memprediksi apakah bahan pangan tersebut mengandung alergen atau tidak, menggunakan algoritma *data mining* SVM. Beberapa keunggulan dari algoritma SVM, yaitu mempunyai akurasi yang tinggi, penggunaan memori yang efisien, dan mampu menangani data terdistribusi tidak normal. Tahapan penelitian yang dilakukan adalah dilakukan *data gathering* (pengumpulan data), dilanjutkan dengan *pre-processing data* yang mencakup *data transformation*, *handling missing value*, dan *copy object*. Tahapan selanjutnya adalah validasi, dimana dalam penelitian ini dilakukan komparasi penggunaan *10-fold cross validation*, *split validation* dengan pembagian *data training* dan *data testing* 90% banding 10%, serta *split validation* dengan perbandingan 80% dan 20%. Setelah *dataset* melalui proses validasi, dilanjutkan dengan implementasi algoritma SVM, dan yang terakhir adalah tahap evaluasi menggunakan *confusion matrix*. Nilai akurasi yang dihasilkan berdasarkan komparasi proses validasi, yaitu pada penggunaan *10-fold cross validation*, SVM mempunyai tingkat akurasi sebesar 97,24%. Pada penggunaan *split validation* perbandingan *data training* dan *data testing* 90% banding 10% menghasilkan nilai akurasi sebesar 97,50%. Sedangkan pada penggunaan *split validation* dengan perbandingan 80% dan 20% diperoleh tingkat akurasi sebesar 98,75%.

Kata kunci: Alergen, Penambangan Data, Validasi K-Fold Cross, Validasi Terpisah, Mesin Vektor Dukungan

1. LATAR BELAKANG

Program Makan Bergizi Gratis yang diinisiasi oleh Kementerian Kesehatan merupakan salah satu strategi dalam upaya membangun generasi sehat, cerdas, dan produktif menuju Indonesia Emas 2045 (Direktorat Jenderal Pendidikan Anak Usia Dini, Pendidikan Dasar,

2024). Namun dalam uji coba pelaksanaannya yang telah dimulai sejak Mei 2024 (Institute for Development of Economics and Finance (INDEF), 2024) mendapatkan tantangan penting terkait dengan alergi makanan, hal ini dikarenakan akan berdampak pada kesehatan sasaran program (Asosiasi Institusi Pendidikan Tinggi Vokasi Gizi Indonesia (AIPVOGI), 2025). Beberapa kasus alergi pada pelaksanaan MBG ini diantaranya terjadi pada Sekolah Dasar Negeri (SDN) Labuan Bajo 2, Manggarai Barat, Nusa Tenggara Timur (NTT), dimana tidak semua siswa mengambil jatah makannya dikarenakan alergi makanan tertentu (Ardin, 2025). Kasus lainnya terjadi Sekolah Menengah Pertama Negeri (SMPN) 13 Surabaya, terdapat siswa yang mempunyai alergi ayam (Miranda, 2025). Kasus alergi makanan pada program MBG juga terjadi di Sekolah Dasar Katolik (SDK) Andaluri, Kecamatan Kota Waingapu, Sumba Timur, Nusa Tenggara Timur (NTT), dimana 29 siswanya mengalami alergi (DetikBali, 2025). Siswa di Sekolah Dasar (SD) Barunawati 2 dan SMPN 61 Palmerah, Jakarta Barat juga mengalami hal serupa (News, 2025).

Alergi makanan merupakan gangguan kesehatan yang muncul diakibatkan oleh respon imun spesifik yang muncul dari paparan makanan (Iglesia et al., 2024). Alergi makanan ini dapat terjadi pada segala kelompok umur, dengan prevalensi yang berbeda antara anak dan dewasa, dimana prevalensi pada anak jika dibandingkan dengan orang dewasa lebih besar (Z. Azizah et al., 2024). Identifikasi alergen, yaitu bahan pangan atau senyawa yang dapat menyebabkan alergi (Direktorat Standardisasi Pangan Olahan, 2019) sebagai pemicu munculnya alergi merupakan hal penting yang dapat menjadi salah satu strategi pencegahan (Sari et al., 2021). Menggunakan *dataset* publik produk pangan dan alergen yang diperoleh dari Kaggle (<https://www.kaggle.com>), penelitian ini bertujuan untuk memprediksi apakah bahan pangan tersebut mengandung alergen atau tidak, menggunakan algoritma *data mining* Support Vector Machines (SVM).

SVM merupakan algoritma klasifikasi yang menggunakan teknik *machine learning* untuk pembuatan batasan yang memisahkan data berdasarkan kelasnya (Amna et al., 2023). Beberapa keunggulan dari algoritma SVM, yaitu mempunyai akurasi yang tinggi, pemakaian memori yang efisien, dan mampu dalam penanganan data terdistribusi tidak normal (Aisah et al., 2023). Berdasarkan keunggulan tersebut, penelitian ini menggunakan algoritma SVM.

Beberapa penelitian terdahulu yang menggunakan algoritma SVM untuk memprediksi kandungan alergen diantaranya dilakukan oleh Assa'ad (2024). Dalam penelitiannya, dilakukan komparasi algoritma *Multi-Layer Perceptron* (MLP), *K-Nearest Neighbors* (K-NN), dan *Support Vector Machine* (SVM), yang ternyata menghasilkan bahwa SVM memiliki performansi terbaik (Assa'ad, 2024). Penelitian selanjutnya dilakukan oleh Nedyalkova et al.,

(2023) yang berhasil membuat model prediksi alergen pada protein tanaman, dimana algoritma SVM mempunyai tingkat akurasi tertinggi diantara algoritma KNN, *Random Forest* (RF), dan *Naive Bayes* (NB), yaitu sebesar 93% (Nedyalkova et al., 2023). Wang *et al.*, (2021) dalam penelitiannya mengkomparasikan beberapa algoritma klasifikasi untuk memprediksi alergenitas pada protein makanan. Hasil penelitiannya menunjukkan bahwa algoritma SVM mempunyai tingkat akurasi prediksi yang cukup bagus, sebesar 74,18% dibandingkan dengan algoritma NB sebesar 72,03%, namun dalam kasus ini masih berada di bawah algoritma K-NN (Wang et al., 2021). Selanjutnya terdapat penelitian yang dilakukan oleh Zhang *et al.*, (2023). Zhang et al., melakukan komparasi beberapa algoritma data mining untuk memprediksi hasil *Oral Food Challenges* (OFC) dikarenakan keterbatasan uji klinis. Hasil penelitiannya menunjukkan bahwa algoritma SVM mempunyai tingkat akurasi yang cukup tinggi dibandingkan dengan algoritma NB dan *Logistic Regression* (LR), namun masih berada di bawah akurasi dari algoritma RF dan LUCCK (Zhang et al., 2023). Shaukat *et al.*, (2024) dalam penelitiannya untuk deteksi dan klasifikasi alergen makanan, membandingkan beberapa algoritma *data mining*. Hasil penelitiannya menunjukkan bahwa algoritma SVM mempunyai tingkat akurasi yang cukup bagus sebesar 93,54%. Namun dalam kasus penelitiannya tersebut, tingkat akurasi algoritma SVM berada di bawah algoritma RF sebesar 96,8% dan K-NN sebesar 95,16% (Shaukat et al., 2024).

Penelitian ini menggunakan metode atau teknik evaluasi *confusion matrix* guna mengukur tingkat akurasi dari algoritma SVM yang diimplementasikan.

2. KAJIAN TEORITIS

Data Mining

Data mining merupakan suatu proses ekstraksi atau penggalian pengetahuan dari suatu *dataset* yang kompleks dan berukuran besar (Rahayu et al., 2024). *Data mining* mempunyai tujuan untuk identifikasi suatu pola atau model, hubungan, ataupun informasi yang mungkin tidak bisa dilihat dengan cara langsung pada suatu kumpulan data, sehingga menghasilkan pengetahuan yang bernilai (Narulita et al., 2023). Dalam prosesnya *data mining* menyertakan berbagai fungsi matematis, teknik statistik, serta kecerdasan buatan guna melakukan analisis data secara sistematis dan otomatis. Adapun hasil dari proses *data mining* dapat dipergunakan dalam pendukung keputusan, perumusan strategi bisnis, peningkatan efisiensi proses operasional, identifikasi suatu tren, dan lain sebagainya. Terdapat beberapa teknik utama dalam *data mining*, yaitu (Rahayu et al., 2024):

Classification (klasifikasi)

Teknik ini melakukan pengelompokan data ke dalam suatu kategori (*class*) tertentu, berdasarkan atribut (label) atau karakteristik tertentu.

Regression (regresi)

Teknik ini digunakan untuk pemodelan hubungan antara variabel dependen dan variabel independen. Selain itu juga dipergunakan untuk prediksi nilai kontinu berdasarkan linieritas (linier atau non-linier) diantara variabel-variabel tersebut.

Clustering (klustering)

Metode ini dipergunakan untuk pengelompokan data dalam suatu *cluster* (kelompok) yang didasarkan pada persamaan karakteristik atau fitur tertentu.

Association (asosiasi)

Metode ini dipergunakan untuk menemukan asosiasi (hubungan) antara variabel (*item*) dalam suatu data.

Support Vector Machines (SVM)

SVM merupakan algoritma klasifikasi yang menggunakan teknik *machine learning* untuk pembuatan batasan yang memisahkan data berdasarkan kelasnya (Amna et al., 2023). SVM mempunyai tujuan utama, yaitu menemukan *hyperplane* untuk pemisahan data *input* dengan margin maksimal yang berasal dari kelas yang berbeda (Narulita & Adi, 2024). Terdapat dua jenis SVM, yaitu linier dan non-linier yang dibedakan atas pendefinisian *hyperplane*.

SVM membutuhkan proses pemilihan fitur (*feature*) *input* yang sesuai untuk mendapatkan hasil terbaik, hal ini dikarenakan SVM rentan terhadap masalah ketinggian angka (*height number problem*). Salah satu kelebihan dari SVM adalah algoritma ini kebal terhadap adanya *overfitting* (kondisi yang terlalu rumit dan terlalu sesuai dengan *data training*) atau dapat dikatakan mempunyai risiko *overfitting* rendah, terlebih saat dipergunakan pada proses normalisasi. Sedangkan kelemahan dari SVM adalah membutuhkan waktu lama untuk proses pembelajaran (*learning*), khususnya pada *dataset* yang mempunyai ukuran sangat besar dan non-linier yang lebih kompleks.

Confusion Matrix

Confusion matrix adalah suatu metode atau teknik yang dipergunakan untuk pengukuran performansi atau kinerja model yang terbentuk dengan cara mengkomparasikan hasil prediksi dengan data sebenarnya atau aktual (Muttaqin et al., 2023). Tabel 1 berikut merupakan presentasi dari *confusion matrix*.

Tabel 1. Confusion Matrix

<i>Predicted Value</i>	<i>Actual Value</i>	
	<i>Positive</i>	<i>Negative</i>
<i>Positive</i>	TP	FP
<i>Negative</i>	FN	TN

Keterangan:

- TP : Data positif dan diprediksi positif
- TF : Data negatif dan diprediksi negatif
- FP : Data negatif dan diprediksi positif
- FN : Data positif dan diprediksi negatif

Sedangkan rumus perhitungan untuk setiap indikator pada confusion matrix ditunjukkan pada persamaan (1) sampai (3).

$$Accuracy = \frac{(TP+TN)}{(TP+FP+TN+FN)} \dots\dots\dots (1)$$

$$Recall = \frac{(TP)}{(TP+FN)} \dots\dots\dots (2)$$

$$Precision = \frac{(TP)}{(TP+FP)} \dots\dots\dots (3)$$

Alergen (*Allergen*)

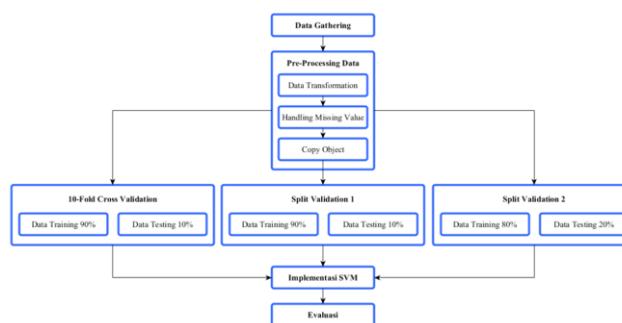
Definisi alergen adalah senyawa atau bahan pangan yang dapat mengakibatkan timbulnya alergi (Direktorat Standardisasi Pangan Olahan, 2019). Alergen itu sendiri dapat berupa bahan atau hasil olahan dari bahan-bahan berikut ini:

- a. Susu (termasuk di dalamnya laktosa);
- b. Telur;
- c. SO₂ dan sulfit dengan jumlah kandungan paling minimal 10 mg/kg, termasuk sebagai SO₂ (natrium bisulfit, kalium bisulfit, natrium metabisulfit, belerang dioksida, kalsium bisulfit, kalium sulfit) untuk produk-produk yang siap konsumsi.
- d. Kedelai dan kacang tanah (*peanut*);
- e. Sereal yang mengandung gluten, seperti *strain* hibrida, *oats*, gandum, *rye*, *spelt*, atau *barley*;

- f. Krustase (kepiting, udang, lobster), ikan, moluska (siput laut, kerang, tiram, atau bekicot);
- g. Kacang pohon (*tree nuts*), termasuk *almond*, kacang *Queensland* atau *Macadamia*, kacang *Brazil*, *walnut*, kacang kenari, *hazelnut*, kacang *pecan*, kacang mede, dan kacang *pistachio*.

3. METODE PENELITIAN

Penerapan algoritma SVM untuk deteksi alergen pada *dataset* produk pangan mempunyai beberapa tahapan proses dalam penyelesaian permasalahan tersebut. Adapun tahapan penelitian ini diperlihatkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

Penjelasan setiap tahapan penelitian pada Gambar 1 dapat dijelaskan sebagai berikut:

Data Gathering (Pengumpulan Data)

Dataset yang dipergunakan di dalam penelitian ini adalah *dataset food ingredients and allergens* dengan *file type* *.CSV. *Dataset* ini merupakan *dataset* publik yang diperoleh dari portal *dataset* untuk *machine learning* Kaggle yang dapat diakses di <https://www.kaggle.com>. *Dataset food ingredients and allergens* terdiri dari tujuh (7) atribut, yaitu *food product*, *main ingredient*, *sweetener*, *fat/oil*, *seasoning*, *allergens*, dan *prediction* sebagai atribut *class* atau *label*. *Dataset* ini mempunyai 399 *record data*. Terdapat data *missing value* pada *dataset* ini, sehingga sebelum implementasi algoritma SVM harus dilakukan pembersihan data (*data cleaning*).

Pre-Processing Data

Pre-processing data merupakan tahap awal penelitian yang dapat melibatkan proses *data integration* (integrasi data), *data labeling* (pelabelan data), *data transformation* (transformasi data), dan juga *data cleaning* (pembersihan data) (Gori et al., 2024). Pada penelitian ini dilakukan beberapa proses terlebih dahulu sebelum *dataset* siap digunakan, yaitu:

a. *Data Transformation* (Transformasi Data)

Data transformation (transformasi data) merupakan tahapan untuk mengubah representasi data sesuai yang dibutuhkan pada penerapan algoritma *data mining* (R. A. Azizah et al., 2022).

b. *Handling Missing Value* (Penanganan *Missing Value*)

Handling missing value (penanganan *missing value*) termasuk dalam bagian *data cleaning* (pembersihan data) yang merupakan serangkaian tahapan yang dilakukan untuk pembersihan data, seperti identifikasi dan pengisian data yang hilang (*missing values*), pencarian dan penghapusan data yang sama (duplikat), penanganan data yang tidak lengkap, serta penghapusan inkonsistensi data yang dapat berupa *noisy* atau *outlier* (Putra & Putri, 2022).

c. *Copy Object*

Tahapan *copy object* yang dimaksud adalah membuat salinan objek pada RapidMiner menggunakan operator *multiply*. Operator ini mengambil objek dari *port input* dan mengirimkan *copy* atau salinannya ke *port output* (Dennis Aprilla C et al., 2013).

1. Validasi

Validasi hasil eksplorasi data merupakan tahap penting dalam suatu proses analisis data untuk memastikan bahwa hasil dari eksplorasi data tersebut andal dan akurat. Validasi hasil eksplorasi data adalah tahap pemeriksaan apakah hasil eksplorasi data sudah mencerminkan data yang sebenarnya (Muttaqin et al., 2023). Pembagian suatu kumpulan data menjadi *data training* dan *data testing* berpengaruh pada nilai akurasi, sehingga kesalahan penentuan proporsi data ini dapat membawa dampak pada nilai akurasi yang dihasilkan (Musu et al., 2021). Pada penelitian ini dilakukan komparasi terhadap beberapa teknik validasi berikut:

a. *10-Fold Cross Validation* (*Data Training* 90% dan *Data Testing* 10%)

K-Fold Cross Validation adalah teknik *re-sampling* untuk penyesuaian parameter data ketika pembangunan model. *Fold* merupakan jumlah himpunan *training* yang akan dihasilkan. Proses ini membagi *dataset* menjadi *data training* dan *data testing* (Irawan et al., 2024). Penentuan nilai K disesuaikan dengan kebutuhan penelitian. Dalam penelitian ini digunakan nilai $K = 10$ untuk mendapatkan akurasi yang optimal.

b. *Split Validation 1* (*Data Training* 90% dan *Data Testing* 10%)

Split validation berfungsi untuk melakukan validasi sederhana secara *random* dan membagi *dataset* menjadi dua (2) bagian, yaitu *data training* dan *data testing* dalam proses evaluasi model. Hal ini dilakukan untuk memperkirakan seberapa akurat model yang dihasilkan (RapidMiner, 2024). Pada tahap *split validation* ini, pembagian *data training* dan *data testing* adalah sebesar 90% dan 10%.

c. *Split Validation 2 (Data Training 80% dan Data Testing 20%)*

Sama halnya dengan tahap *split validation* sebelumnya, hanya saja pada *split validation* kedua ini digunakan pembagian *data training* dan *data testing* adalah sebesar 80% dan 20%.

2. Implementasi SVM

Setelah tahap validasi selesai dilakukan, tahap selanjutnya adalah tahap penerapan atau implementasi algoritma SVM. Pada *tool* RapidMiner penggunaan operator *apply model* mempunyai kemampuan untuk mengimplementasikan SVM ke data baru. Operator ini mempergunakan model *training* untuk melakukan prediksi nilai atribut keluaran untuk data yang baru. Sedangkan operator kinerja (*performance*) berfungsi untuk evaluasi performansi atau kinerja dari algoritma SVM (Aisah et al., 2023).

3. Evaluasi

Penelitian ini menggunakan metode evaluasi *confusion matrix* untuk mengukur performansi dari algoritma SVM dengan cara melakukan perbandingan terhadap hasil prediksi dengan data aktual (Theet et al., 2010).

4. HASIL DAN PEMBAHASAN

Pada tahap *data gathering* atau pengumpulan data diperoleh *dataset food ingredients and allergens* dengan *file type* *.CSV. *Dataset* ini merupakan *dataset* publik yang diperoleh dari portal *dataset* untuk *machine learning* Kaggle yang dapat diakses di <https://www.kaggle.com>. *Dataset food ingredients and allergens* terdiri dari tujuh (7) atribut, yaitu *food product*, *main ingredient*, *sweetener*, *fat/oil*, *seasoning*, *allergens*, dan *prediction* sebagai atribut *class* atau *label*. *Dataset* ini mempunyai 399 *record data*. Namun, pada *dataset* ini terdapat satu (1) data *missing value*, sehingga perlu dilakukan pembersihan data (*data cleaning*).

Tahap *pre-processing* data pada penelitian ini dilakukan beberapa sub tahapan, yang pertama *data transformation* (transformasi data). Dikarenakan terdapat data *missing value* pada

dataset yang digunakan, yang mana nilai ini diisi dengan nilai rata-rata (*average*) data sekitarnya, maka diperlukan proses transformasi data dari nominal (data kategorik) ke numerik. Setelah dilakukan transformasi data, kemudian dilakukan *handling missing value* atau penanganan *missing value* pada *dataset*. *Data missing value* pada *dataset food ingredients and allergens* berada pada atribut *prediction* sebagai label atau *class*. Penanganan data *missing value* pada RapidMiner menggunakan operator *replace missing value*. Setelah data *missing value* tertangani, dilakukan proses *copy object*, yaitu membuat salinan *dataset* yang akan digunakan pada beberapa teknik validasi melalui penggunaan operator *multiply*.

Pada tahap selanjutnya, yaitu tahap validasi dilakukan sebanyak tiga (3) kali percobaan menggunakan teknik dan pembagian persentase *data training* dan *data testing* berbeda. Validasi pertama menggunakan teknik *10-fold cross validation* dengan pembagian *data training* dan *data testing* masing-masing sebesar 90% dan 10%. Dilanjutkan dengan implementasi algoritma SVM. Pada RapidMiner penggunaan operator *apply model* berfungsi untuk menerapkan model yang terbentuk pada data yang baru, sedangkan untuk melakukan evaluasi model yang dihasilkan digunakan operator *performance*. Hasil dari penggunaan operator *performance* berupa metrik kinerja *accuracy*, *recall*, dan *precision*. Hasil evaluasi algoritma SVM dengan teknik validasi *10-fold cross validation* (persentase *data training* dan *data testing* sebesar 90% dan 10%) ditunjukkan pada Tabel 2.

Tabel 2. Performa SVM dengan Validasi 10-Fold Cross Validation

<i>accuracy: 97.24% ± 2.19% (micro average: 97.24%)</i>			
	<i>true Contains</i>	<i>true Does not contain</i>	<i>class precision</i>
<i>pred. Contains</i>	247	2	99.20%
<i>pred. Does not contain</i>	9	141	94.00%
<i>class recall</i>	96.48%	98.60%	

Pada percobaan kedua, penerapan algoritma SVM dengan teknik validasi *split validation* dan pembagian *data training* dan *data testing* masing-masing sebesar 90% dan 10%, hasil evaluasinya ditunjukkan pada Tabel 3.

Tabel 3. Performa SVM dengan Split Validation (90%:10%)

<i>accuracy: 97.50%</i>			
	<i>true Contains</i>	<i>true Does not contain</i>	<i>class precision</i>
<i>pred. Contains</i>	25	0	100.00%
<i>pred. Does not contain</i>	1	14	93.33%

<i>class recall</i>	96.15%	100.00%
---------------------	--------	---------

Sedangkan untuk percobaan ketiga, penerapan algoritma SVM dengan teknik validasi *split validation* dan pembagian *data training* dan *data testing* masing-masing sebesar 80% dan 20%, hasil evaluasinya ditunjukkan pada Tabel 4.

Tabel 4. Performa SVM dengan *Split Validation* (80%:20%)

<i>accuracy: 98.75%</i>			
	<i>true Contains</i>	<i>true Does not contain</i>	<i>class precision</i>
<i>pred. Contains</i>	50	0	100.00%
<i>pred. Does not contain</i>	1	29	96.67%
<i>class recall</i>	98.04%	100.00%	

Berdasarkan tabel di atas, dapat dirangkum hasil evaluasi dari beberapa percobaan penerapan algoritma SVM dengan teknik validasi yang dapat dilihat pada Tabel 5 di bawah.

Tabel 5. Perbandingan Hasil Evaluasi (*Confusion Matrix*)

Percobaan	Accuracy	Recall	Precision
SVM + 10-Fold Cross Validation	97,24%	96,48	99,20%
SVM + <i>Split Validation</i> (<i>Data Training</i> 90%, <i>Data Testing</i> 10%)	97,50%	96,15%	100,00%
SVM + <i>Split Validation</i> (<i>Data Training</i> 80%, <i>Data Testing</i> 20%)	98,75%	98,04%	100,00%

Dari Tabel 5 terlihat bahwa implementasi algoritma SVM dengan teknik *split validation* dengan pembagian *data training* sebesar 80% dan *data testing* 20% menghasilkan nilai *accuracy* 98,75%. Nilai *recall* sebesar 98,04% menunjukkan bahwa algoritma SVM mampu memprediksi keseluruhan data yang diprediksi positif, sedangkan tingkat *precision* sebesar 100,00% menunjukkan bahwa algoritma SVM mempunyai kemampuan untuk memprediksi secara tepat data yang benar-benar positif.

KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian pada *dataset food ingredients and allergens* melalui perbandingan tiga (3) teknik validasi, yaitu 10-fold cross validation dengan pembagian persentase data training dan data testing masing-masing sebesar 90% dan 10%, *split validation*

dengan pembagian persentase data training dan data testing masing-masing sebesar 90% dan 10%, serta split validation dengan pembagian persentase data training dan data testing masing-masing sebesar 80% dan 20%, diperoleh hasil evaluasi bahwa penerapan algoritma SVM dengan teknik split validation (pembagian persentase *data training* 80% dan *data testing* 20%) mempunyai performansi paling baik, yaitu nilai *accuracy* sebesar 98,75%, *recall* sebesar 98,04%, dan *precision* sebesar 100%. Dari sini dapat disimpulkan bahwa penentuan teknik validasi dan pembagian persentase *data training* dan *data testing* dapat memberikan hasil (*accuracy*, *recall*, dan *precision*) yang berbeda.

Pada penelitian selanjutnya dapat dilakukan penerapan algoritma *data mining* lainnya, seperti Decision Tree, Naive Bayes, K-NN, Neural Networks (NN), Logistic Regression, atau Convolutional Neural Network (CNN) pada *dataset food ingredients and allergens*. Selain itu direkomendasikan penggunaan *feature selection*, seperti *Information Gain*, *Chi-Square*, *Correlation-based Feature Selection*, *Recursive Feature Elimination (RFE)*, *Forward Selection*, *Backward Elimination*, atau Regularisasi (*Lasso* dan *Ridge*) untuk menemukan model yang paling efektif dalam meningkatkan performansi algoritma SVM.

DAFTAR REFERENSI

- Aisah, I. S., Irawan, B., & Suprpti, T. (2023). Algoritma Support Vector Machine (SVM) untuk analisis sentimen ulasan aplikasi Al Qur'an digital. *JATI: Jurnal Mahasiswa Teknik Informatika*, 7(6), 3759–3765. <https://doi.org/10.36040/jati.v7i6.8263>
- Amna, S. W., Sudipa, I. G. I., Putra, T. A. E., Wahidin, A. J., Syukrilla, W. A., Wardhani, A. K., Heryana, N., Indriyani, T., & Santoso, L. W. (2023). *Data mining*. PT Global Eksekutif Teknologi.
- Ardin, A. (2025, February 17). Tak semua siswa ambil jatah MBG di Labuan Bajo karena alergi-vegetarian. *Detik*. <https://www.detik.com>
- Asosiasi Institusi Pendidikan Tinggi Vokasi Gizi Indonesia (AIPVOGI). (2025, January 6). Standar gizi dan makanan dalam program makan bergizi gratis tahun 2025. <https://aipvogi.org>
- Assa'ad, A. H. (2024). Implementasi machine learning dengan metode KNN, SVM, dan MLP dalam mendeteksi alergen makanan pada resep makanan [Institut Teknologi Sepuluh Nopember]. <http://repository.its.ac.id/id/eprint/110334>
- Azizah, R. A., Bachtiar, F., & Adinugroho, S. (2022). Klasifikasi kinerja akademik siswa menggunakan neighbor weighted K-nearest neighbor dengan seleksi fitur information gain. *JTIK: Jurnal Teknologi Informasi Dan Ilmu Komputer*, 9(3), 605–614. <https://doi.org/10.25126/jtiik.2022935751>

- Azizah, Z., Falihah, A. H., Santoso, B. B. A., Puspitasari, I., & Sahid, M. N. A. (2024). Frequency of food allergy based on a survey of adults in Yogyakarta and Java regions. *Majalah Farmaseutik*, 20(2), 276–281. <https://doi.org/10.22146/farmaseutik.v20i2.85546>
- Dennis Aprilla C., Donny Aji Baskoro, Ambarwati, L., & Wicaksana, I. W. S. (2013). Belajar data mining dengan Rapid Miner (R. Sanjaya, Ed.). Open Content Model.
- DetikBali, T. (2025, February 22). Pengelola dapur bantah 29 siswa di Sumba Timur keracunan setelah santap MBG. Detik. <https://www.detik.com>
- Direktorat Jenderal Pendidikan Anak Usia Dini, Pendidikan Dasar, dan Pendidikan Menengah (2024). Pedoman makan bergizi gratis (MBG) di satuan pendidikan. Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. <https://aipvogi.org>
- Direktorat Standardisasi Pangan Olahan. (2019). Pedoman implementasi pelabelan pangan olahan. Deputi Bidang Pengawasan Pangan Olahan, Badan Pengawas Obat dan Makanan Republik Indonesia.
- Gori, T., Sunyoto, A., & Fatta, H. Al. (2024). Preprocessing data dan klasifikasi untuk prediksi kinerja akademik siswa. *JTIK: Jurnal Teknologi Informasi Dan Ilmu Komputer*, 11(1), 215–224. <https://doi.org/10.25126/jtiik.20241118074>
- Iglesia, E. G. A., Kwan, M., Virkud, Y. V., & Iweala, O. I. (2024). Management of food allergies and food-related anaphylaxis. *JAMA*, 331(6), 510–521. <https://doi.org/10.1001/jama.2023.26857>
- Institute for Development of Economics and Finance (INDEF). (2024). Efek pengganda program makan bergizi gratis. <https://indef.or.id>
- Irawan, R. N., Hindrayani, K. M., & Idhom, M. (2024). Penerapan cross validation sebagai analisis sentimen pelayanan publik kereta api lokal Daop 8 menggunakan metode multinomial naive bayes. *G-Tech: Jurnal Teknologi Terapan*, 8(2), 954–963. <https://doi.org/10.33379/gtech.v8i2.4117>
- Miranda, S. (2025, January 13). Hari pertama program MBG di Surabaya, alergi menu tertentu jadi catatan khusus. *Ketik.co.id*. <https://ketik.co.id>
- Musu, W., Ibrahim, A., & Heriadi. (2021). Pengaruh komposisi data training dan testing terhadap akurasi algoritma C4.5. *SISITI: Seminar Sistem Informasi Dan Teknologi Informasi*, 10(1), 186–195. <https://ejurnal.undipa.ac.id/index.php/sisiti/article/view/802>
- Muttaqin, W., Widiyanto, W. W., Munsarif, M., Mandias, G. F., Pungus, S. R., Widarman, A., Hapsari, W. K., Hardiyanti, S. A., Fatkhudin, A., Pasnur, B., Anshori, M., Suryani, & Saputra, N. (2023). Pengenalan data mining. Yayasan Kita Menulis.
- Narulita, S., & Adi, P. N. (2024). Feature selection information gain pada klasifikasi pasien penyakit jantung (heart disease). *JURMIK: Jurnal Rekam Medis Dan Manajemen Informasi Kesehatan*, 4(1), 13–19. <https://doi.org/10.53416/jurmik.v4i1.240>

- Narulita, S., Prihati, O., Oktaga, A. T., & Widyantoro, A. E. (2023). Performansi algoritma clustering K-means untuk penentuan status malnutrisi pada balita. *Jurnal Informasi, Sains, Dan Teknologi*, 6(1), 188–202. <https://isaintek.polinef.ac.id/index.php/isaintek/article/view/128>
- Nedyalkova, M., Vasighi, M., Azmoon, A., Naneva, L., & Simeonov, V. (2023). Sequence-based prediction of plant allergenic proteins: Machine learning classification approach. *ACS Omega*, 8(4), 3698–3704. <https://doi.org/10.1021/acsomega.2c02842>
- News, T. (2025, January 12). Seorang pelajar tak bisa makan menu MBG karena alasan alergi, begini respons Cak Imin. *Liputan6.com*. <https://www.liputan6.com>
- Putra, M. Y., & Putri, D. I. (2022). Pemanfaatan algoritma naive bayes dan K-nearest neighbor untuk klasifikasi jurusan siswa kelas XI. *Jurnal Tekno Kompak*, 16(2), 176–187. <https://doi.org/10.33365/jtk.v16i2.2002>
- Rahayu, P. W., Sudipa, I. G. I., Suryani, S., Surachman, A., Ridwan, A., Darmawiguna, I. G. M., Sutoyo, M. N., Slamet, I., Harlina, S., & Maysanjaya, I. M. D. (2024). Buku ajar data mining. PT Sonpedia Publishing Indonesia. https://www.researchgate.net/publication/377415198_BUKU_AJAR_DATA_MINING
- RapidMiner, I. (2024). Split validation (RapidMiner Studio Core). <https://docs.rapidminer.com>
- Sari, K., Palupi, N. S., & Giriwono, P. E. (2021). Allergen sanitation in biscuit production process to reduce egg allergen residu. *Journal of Food Technology and Industry*, 32(2), 136–147. <https://doi.org/10.6066/jtip.2021.32.2.136>
- Shaukat, H., Sultan, A., & Salahuddin, H. (2024). Enhancing food safety: A machine learning approach for accurate detection and classification of food allergens. *Journal of Computing & Biomedical Informatics, Special Is*, 1–16. <https://jcbi.org/index.php/Main/article/view/474/371>
- Thet, T. T., Na, J.-C., & Khoo, C. S. G. (2010). Aspect-based sentiment analysis of movie reviews on discussion boards. *Sage Journals*, 36(6), 823–848. <https://doi.org/10.1177/0165551510388123>
- Wang, L., Niu, D., Zhao, X., Wang, X., Hao, M., & Che, H. (2021). A comparative analysis of novel deep learning and ensemble learning models to predict the allergenicity of food proteins. *Foods*, 10(809), 1–15. <https://doi.org/10.3390/foods10040809>
- Zhang, J., Lee, D., Jungles, K., Shaltis, D., Najarian, K., Ravikumar, R., Sanders, G., & Gryak, J. (2023). Prediction of oral food challenge outcomes via ensemble learning. *Informatics in Medicine Unlocked*, 36, 1–10. <https://doi.org/10.1016/j.imu.2022.101142>