

Pencarian Bahan Magnet Kuat Dua Dimensi menggunakan *Support Vector Machine*

Muchamad Fajar Alif¹, Nurul Ikhsan²

^{1,2}Fakultas Informatika, Universitas Telkom, Bandung

¹muchamadfajaralif@students.telkomuniversity.ac.id, ²ikhsan@telkomuniversity.ac.id

Abstrak

Penemuan dan desain material baru merupakan tujuan utama dari ilmu material. Masih tingginya jumlah data material yang perlu di analisis untuk mencari bahan baru menjadi tantangan di bidang ini karena sangat memakan waktu dan sumber daya komputasi jika semua kemungkinan struktur dari banyak kombinasi material dihitung dengan menggunakan metode yang presisi, seperti *Density Functional Theory* (DFT). Menghadapi tantangan tersebut, dibutuhkan suatu sistem yang dapat menganalisis banyaknya kombinasi struktur material dengan kinerja cepat namun tetap mempertahankan presisi dari perhitungan. Disinilah *materials informatics* hadir untuk mengakselerasi proses pencarian bahan unggul yang baru, *material informatics* merupakan suatu bidang keilmuan baru yang bertujuan untuk memanfaatkan ilmu informatika dalam metode komputasi yang umum digunakan khususnya pada ilmu material. Pada penelitian ini, 198 kombinasi bahan magnet dua dimensi struktur $A_2B_2X_6$ akan diklasifikasikan menggunakan *support vector machine* (SVM). Dengan menerapkan seleksi fitur, model klasifikasi SVM yang dibangun dapat menghasilkan nilai akurasi dan *f1-score* sekitar 97%.

Kata kunci : material science, density functional theory, material informatics, machine learning, support vector machine

Abstract

The discovery and design of new materials is the main goal of material science. The large amount of material data that needs to be analyzed to find new materials is a grand challenge in this field because it is taking very high computational time and resources if we calculate all of the possible structure from many material combinations using precision methods, such as *Density Functional Theory* (DFT). Facing these challenges, we need a system that can analyze the combination of material structures with fast performance but still maintain the precision of calculation. This is where *materials informatics* is present to accelerate the search for new superior material, *material informatics* is a new scientific field which aims to utilize information in the computational methods commonly used in scientific material. In this research, 198 combinations of two-dimensional magnetic materials of $A_2B_2X_6$ structure will be classified using *support vector machine* (SVM). By applying feature selection, the SVM classification model that is built can produce an accuracy and an *f1-score* of around 97%.

Keywords: material science, density functional theory, material informatics, machine learning, support vector machine

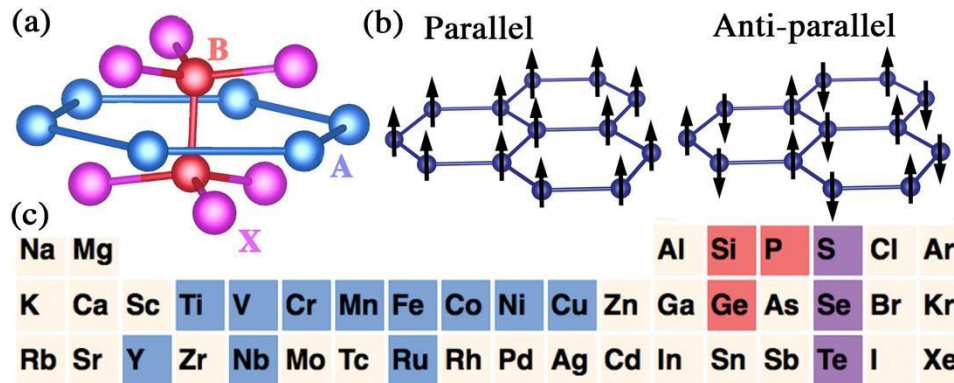
1. Pendahuluan

Latar Belakang

Pencarian bahan alternatif untuk magnet kuat menggunakan simulasi komputer berbasis *Density Functional Theory* (DFT) sudah sangat populer digunakan, metode ini dikenal akurat memprediksi karakter suatu bahan tanpa perlu informasi semi-empiris dari percobaan. Keberhasilan DFT dalam memprediksi suatu bahan diiringi dengan meningkatnya kemampuan komputasi beberapa dekade terakhir telah meningkatkan kuantitas dan kompleksitas data hasil eksperimen yang disediakan. Sejumlah besar data yang dihasilkan ini perlu disimpan dan ditafsirkan guna mendukung penelitian dalam ilmu material. Namun, salah satu keterbatasan yang dimiliki DFT adalah harus mempertimbangkan interaksi dari banyak elektron dan elektron-inti dalam suatu bahan. Karena keterbatasan ini, DFT akan memakan banyak waktu, dan sumber daya sehingga diperlukan sistem yang dapat mengakselerasi proses pencarian bahan.

Untuk dapat menghasilkan wawasan dari basis data material yang sudah tersedia sebagai acuan informasi, para peneliti ilmu material mulai menerapkan pendekatan baru yang dapat digunakan untuk menganalisa data secara otomatis dan berkecepatan tinggi. Terinspirasi dari *data science*, untuk menghadapi tantangan ini, peneliti ilmu material telah berusaha memanfaatkan pembelajaran mesin serta kecerdasan buatan dengan cara mengenali pola, dan korelasi yang terdapat pada data yang tersedia. Hal ini dapat dilakukan untuk membentuk dasar untuk memahami perilaku material serta memprediksi bahan unggul. Pembelajaran mesin dapat mengidentifikasi hubungan dalam data yang kompleks, bahkan untuk data berdimensi tinggi, sehingga waktu untuk menganalisa material dapat dipangkas dengan sangat signifikan [1] dibanding dengan menggunakan DFT. Penelitian yang

dilakukan akan membahas lebih dalam mengenai klasifikasi bahan magnet dua dimensi dari bentuk $A_2B_2X_6$ berdasarkan bahan *perovskite crystal* $Cr_2Ge_2Te_6$ (Gambar 1.(a)) dengan menggunakan 132 dari 198 struktur (Gambar 1.(c)) yang sudah pernah digunakan oleh Rhone. et al. [1].



Gambar 1. Kombinasi Bahan $Cr_2Ge_2Te_6$

Pada penelitian ini, struktur $X = S$ akan diklasifikasikan berdasarkan posisi atom, momen magnetik, jumlah elektron dalam sistem, dan total energi dengan memanfaatkan informasi dari struktur $X = Te, Se$ sebagai acuan. Pendekatan seperti *kernel ridge regression*, *extra tree regression* dan *neural network* telah menunjukkan hasil yang cukup memuaskan pada bidang keilmuan *material informatics* [1, 8]. Pada penelitian ini akan dibangun sebuah sistem untuk mengklasifikasikan bahan dua dimensi untuk mencari magnet unggul memanfaatkan pembelajaran mesin menggunakan *support vector machine* (SVM). Adapun alasan SVM digunakan adalah karena kemampuannya untuk melakukan klasifikasi dua label maupun multi-label, dan juga sangat cocok untuk digunakan dengan ukuran data yang relatif kecil, SVM juga telah menunjukkan hasil yang cukup memuaskan ketika bekerja dengan desain material [10].

Topik dan Batasannya

Batasan masalah dari penelitian ini adalah sebagai berikut:

1. Dari total 198 konfigurasi yang mungkin, hanya 132 kombinasi bahan yang akan digunakan pada proses pembelajaran mesin penelitian ini
2. Sisa konfigurasi yang tersedia akan digunakan untuk keperluan eksplorasi dan pengujian sistem yang dibangun kedepannya

Tujuan

Tujuan yang diharapkan tercapai dari penelitian ini adalah sebagai berikut:

1. Menggunakan metode alternatif untuk mengakselerasi proses pencarian bahan magnet kuat
2. Membangun sistem pencari bahan magnet kuat dengan menggunakan klasifikasi SVM
3. Pengukuran kinerja model klasifikasi SVM yang dibangun

Organisasi Tulisan

Makalah ini disusun dalam 4 bagian. Bagian 2 menjelaskan penelitian-penelitian terkait yang pernah dilakukan untuk menyelesaikan masalah yang sama. Pada Bagian 3 akan dijelaskan gambaran umum sistem yang akan dibuat, mulai dari metode pembuatan dataset, pengujian, serta evaluasi sistem. Bagian 4 menyajikan hasil dari eksperimen dan pengujian yang dilakukan selama penelitian. Terakhir, Bagian 5 menjelaskan kesimpulan dari metode yang diusulkan untuk memecahkan masalah serta rencana pengembangan kedepannya.

2. Studi Terkait

2.1. Materials Informatics

Ide untuk mengakselerasi pencarian material diinisiasi dengan didirikannya *The Materials Genome Initiative* [2] pada tahun 2011 yang merupakan gerakan kolaborasi antara ilmuwan material dengan ilmuwan komputer untuk menggunakan metodologi komputasi untuk memprediksi, menyaring (*screening*), dan mengakselerasi proses pencarian serta analisis material. Berangkat dari paradigma yang sebelumnya digunakan seperti eksperimen, teori, serta simulasi komputer, dalam beberapa dekade terakhir penggunaan metode simulasi *Density Functional Theory* (DFT) untuk menafsirkan dan memprediksi karakteristik material telah berkembang diiringi dengan sumber daya komputasi yang sudah tersedia secara luas dengan diperkenalkannya teknik baru seperti *High Throughput* (HT) [4]. Saat ini telah tersedia basis data material yang berisi hasil puluhan ribu karakteristik material yang

dihasilkan oleh kalkulasi DFT [2] yang sudah dapat diakses oleh para peneliti yang dapat mendorong pendekatan kolaboratif baru untuk penemuan material.

Dengan memasuki paradigma baru yaitu penelitian berbasis data, untuk mengekstrak wawasan dari basis data material yang telah tersedia [12] dengan kinerja yang optimal, peneliti ilmu material sudah mulai mengadopsi teknik *soft computing* untuk memecahkan masalah ini. Makalah tentang desain material yang ditulis oleh Lu. et al. [10], memanfaatkan *support vector machine* (SVM) untuk mengklasifikasikan struktur ABX_3 . Rhone. et al. [1] melakukan prediksi sifat magnetik serta *formation energy* dari 4,223 material struktur $A_2B_2X_6$ berdasarkan bahan $Cr_2Ge_2Te_6$ menggunakan *kernel ridge regression*, *extra tree regression*, dan *neural network*. Tawfik. et al. [3] menggunakan *neural network*, SVM, dan *random forest*, untuk memprediksi sifat struktural dan elektronik dari sekitar 1,500 struktur hanya dengan memanfaatkan 267 data dari hasil simulasi DFT.

2.2. Density Functional Theory

Density Functional Theory (DFT) [6, 7] merupakan suatu metode yang menyelesaikan permasalahan elektron banyak pada bahan (material) dengan memanfaatkan fungsional kerapatan elektron untuk menemukan solusi numerik yang mendekati solusi eksak yang didasari oleh fungsi aproksimasi Born – Oppenheimer.

$$H\psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = E\psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$$

(1)

Sehingga, dengan memperoleh *ground-state function* melalui bantuan simulasi (komputasi) kita bisa memprediksi karakteristik fisis dari suatu bahan, tanpa bantuan parameter apapun dari hasil eksperimen (*ab-initio*). Metode ini merupakan model pendekatan yang paling populer digunakan untuk menghitung sifat maupun struktur dari suatu bahan/material. Dalam ilmu material, DFT dapat digunakan untuk menggambarkan perilaku material dengan memperhatikan kerapatan dan interaksi elektron-elektron yang ada di dalam suatu sistem. DFT menggunakan struktur kristal dari bahan seperti posisi atom, sebagai masukan dan dapat mengeluarkan kerapatan muatan elektron, konfigurasi magnetik, atau energi total sebagai hasilnya. Salah satu contoh penggunaan DFT terhadap suatu bahan yang telah dikonfirmasi oleh eksperimen adalah penelitian yang dilakukan oleh Nozaki. et al. [9] yang menunjukkan bahwa doping iridium skala atom dalam lapisan Fe sangat efektif digunakan dalam *magnetic tunnel junction* (MTJ) berbasis Fe/MgO menggunakan *first principle calculations*.

2.3. Support Vector Machine

Support vector machine (SVM) merupakan salah satu metode atau teknik pembelajaran mesin model *supervised learning* yang sangat populer digunakan untuk masalah klasifikasi, dan regresi. SVM pertama kali diperkenalkan oleh Vapnik [5] dan pada awalnya digunakan untuk klasifikasi data ke dalam dua kelas. SVM bekerja dengan cara memisahkan himpunan data berdasarkan kelasnya secara linier menggunakan *hyperplane* (batas keputusan) sebagai pemisah. Pada perkembangannya, SVM dapat bekerja dengan data multi-kelas, dan dapat digunakan untuk masalah regresi [10]. SVM bekerja dengan berusaha menemukan *hyperplane* optimal yang dapat direpresentasikan dengan persamaan (2).

$$\mathbf{w} \cdot \mathbf{x} + b = 0 \quad (2)$$

Berdasarkan persamaan (2), \mathbf{w} merupakan weight vector, \mathbf{x} merupakan vektor untuk set fitur, dan

b merupakan bias. *Hyperplane* optimal adalah yang memiliki jarak atau *margin* terbesar antara *support vector*. *Support vector* adalah titik data yang terletak paling dekat ke permukaan keputusan (*hyperplane*) atau data terluar pada setiap kelas. SVM mencoba membuat *hyperplane* tepat ditengah-tengah kedua kelas, dengan asumsi terdapat dua kelas yaitu positif dan negatif, sehingga pemisahan antara dua kelas menjadi selebar mungkin. Dengan demikian, *margin* yang ingin didapatkan adalah yang memiliki jarak paling besar dari *hyperplane* kelas positif yang didefinisikan dengan persamaan (3) dengan *hyperplane* kelas negatif yang didefinisikan dengan persamaan (4).

$$\mathbf{w} \cdot \mathbf{x} + b = 1 \quad (3)$$

$$\mathbf{w} \cdot \mathbf{x} + b = -1 \quad (4)$$

Dengan demikian, pencarian *hyperplane* optimal dengan memaksimalkan *margin* yang dapat ditulis

dengan $\frac{-2}{\|\diamond\|}$, atau $\frac{1}{\|\diamond\|^2}$. Memaksimalkan *margin*, dapat dilakukan dengan meminimasi $\|\diamond\|$, yang $\|\diamond\|^2$ kemudian, hal ini akan mengarahkan kita kepada masalah optimasi

$$\min_{\phi_i, b} \frac{1}{2} \|\phi_i\|^2 \tag{5}$$

dengan batasan

$$\phi_i(\phi_i \cdot \phi_i + b) - 1 \geq 1, i = 1, \dots, n \tag{6}$$

dimana ϕ_i merupakan kelas atau keluaran dari SVM. $\phi_i \in \{-1, 1\}$ dengan asumsi jumlah kelas adalah dua, yaitu positif dan negatif, dan i adalah nomor data ke- i pada dataset. Pencarian *hyperplane* optimal merupakan masalah *Quadratic Programming* (QP), masalah ini dapat diselesaikan dengan menggunakan *Lagrange Multiplier*.

Formulasi diatas digunakan dengan berasumsi bahwa data dapat dipisahkan secara linier. Pada umumnya himpunan data tidak selalu terpisah secara linier dan membutuhkan *hyperplane* yang bukan berbentuk garis lurus. Untuk data nonlinier, formulasi diatas akan sedikit berubah dengan menambahkan sebuah variabel kendur atau *slack variable* ξ dengan nilai lebih dari 0 sehingga formulasi diatas menjadi

$$\min_{\phi_i, b} \frac{1}{2} \|\phi_i\|^2 + C \sum_i \xi_i \tag{7}$$

dengan batasan

$$\phi_i(\phi_i \cdot \phi_i + b) - 1 \geq 1 - \xi_i$$

$$(8)$$

Parameter C digunakan untuk mengontrol optimasi antara *margin* dengan kesalahan klasifikasi ξ_i . Dengan demikian, semakin besar C maka semakin besar juga penalti kesalahan klasifikasi. Untuk mendapatkan *hyperplane* yang optimal untuk data nonlinier, *kernel trick* pada SVM dapat diterapkan. Ide dari *kernel trick* adalah dengan memetakan himpunan data ϕ_i ke ruang vektor yang berdimensi lebih tinggi dengan sebuah fungsi transformasi $\psi(\phi_i)$. Pada umumnya transformasi ψ tidak diketahui, dan sangat sulit untuk dipahami, oleh karena itu tidak perlu melakukan transformasi secara eksplisit, sehingga, dapat digantikan dengan kalkulasi *dot product* yang didefinisikan dengan fungsi kernel $\kappa(\phi_i, \phi_j)$. Gagasan untuk tidak pernah secara eksplisit mengubah vektor ϕ_i ke ruang dimensi yang lebih tinggi disebut *kernel trick*. Fungsi *kernel* yang umum digunakan ada empat dan ditunjukkan pada Tabel (1).

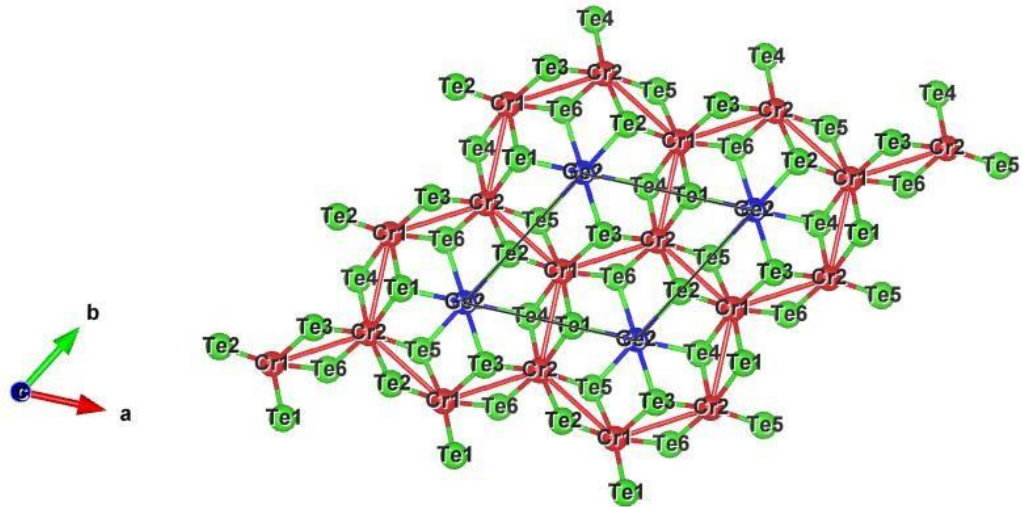
Tabel 1. Kernel function

<i>Kernel</i>	<i>Kernel Function</i>	<i>Parameter</i>
<i>Linear</i>	$\kappa(\phi_i, \phi_j) = \phi_i \cdot \phi_j$	C
<i>Gaussian / Radial Basis Function (RBF)</i>	$\kappa(\phi_i, \phi_j) = \exp(-\gamma \ \phi_i - \phi_j\ ^2)$	C, γ
<i>Polynomial</i>	$\kappa(\phi_i, \phi_j) = \epsilon \kappa(\phi_i, \phi_j) (-\gamma \ \phi_i - \phi_j\ ^2)$	C, d

3. Sistem yang Dibangun

3.1. Struktur Kristal

Berikut adalah struktur kristal dari bahan $Cr_2Ge_2Te_6$ yang akan digunakan untuk membangun dataset, setiap monolayer dibentuk oleh $CrTe$ *octahedra*, dan pasangan Ge terletak tepat di tengah sarang lebah *octahedra* [11]. Kalkulasi DFT akan dilakukan dengan melakukan substitusi atom-atom Cr (situs A) dari daftar: Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Y, Nb, Ru. Dua situs B akan diganti dengan kombinasi atom Ge, Si, dan P, yaitu Ge_2 , $GeSi$, GeP , Si_2 , SiP , P_2 . Enam situs X adalah S, Se, atau Te, yaitu S_6 , Se_6 , dan Te_6 .



Gambar 2. Struktur kristal Cr₂Ge₂Te₆

3.2. Dataset

Dataset yang akan digunakan pada penelitian ini adalah 132 struktur X = Te (Tabel 1), dan X = Se (Tabel 2) sebagai set data latih, lalu untuk 66 struktur X = S akan digunakan sebagai set data uji. Setiap konfigurasi bahan memiliki 45 fitur yang terdiri dari momen magnetik dari setiap struktur (Tabel 3), posisi atom (Tabel 4), jumlah elektron dalam sistem, energi fermi, energi band, serta total energi yang merupakan hasil dari kalkulasi DFT dengan menggunakan *Vienna ab initio Simulation Package* (VASP).

Tabel 2. Konfigurasi bahan Te

	Te					
Ti	1	6
V
Cr
Mn
Fe
Co
Ni
Cu
Y
Nb
Ru	61	66
	Si	Si	Si	Ge	P	P
	Si	Ge	P	Ge	Ge	P

Tabel 3. Konfigurasi bahan Se

	Se					
Ti	67	72
V
Cr
Mn
Fe
Co
Ni
Cu
Y
Nb
Ru	127	132
	Si	Si	Si	Ge	P	P
	Si	Ge	P	Ge	Ge	P

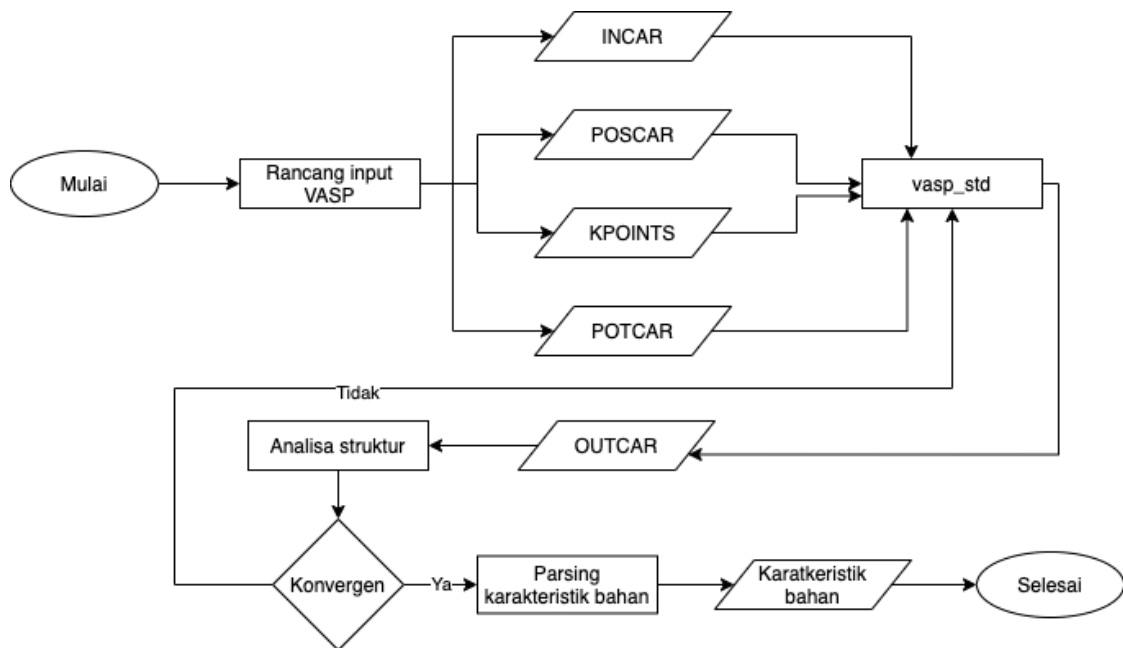
Tabel 4. Momen magnetik

Atom Ke	μ_{total}
1	μ_{atom_1}
2	μ_{atom_2}
3	μ_{atom_3}
...	...
...	...
...	...
...	...
...	...
10	$\mu_{atom_{10}}$
Total	μ_{total}

Tabel 5. Posisi atom

Atom Ke	X	Y	Z
1	X_1	Y_1	Z_1
2	X_2	Y_2	Z_3
3	X_3	Y_2	Z_3
...
...
...
...
...
10	X_{10}	Y_{10}	Z_{10}

Pembuatan dataset menggunakan program VASP dijalankan pada server dengan spesifikasi *processor* Intel® Xeon® CPU E5-2620 v4 @ 2.10 Ghz dengan jumlah *core* sebanyak 16, dan memori sebesar 32 GB. Untuk menjalankan perhitungan terhadap suatu kombinasi, diperlukan empat file masukkan yaitu INCAR, POSCAR, POTCAR dan KPOINTS. File INCAR berisi parameter-parameter kalkulasi yang digunakan untuk simulasi, POSCAR berisi posisi serta konfigurasi setiap atom, POTCAR berisi informasi relevan mengenai *pseudopotential* yang diperlukan untuk melakukan perhitungan dan sudah disediakan oleh VASP, lalu yang terakhir KPOINTS berisi *k-points mesh* yang digunakan untuk perhitungan.



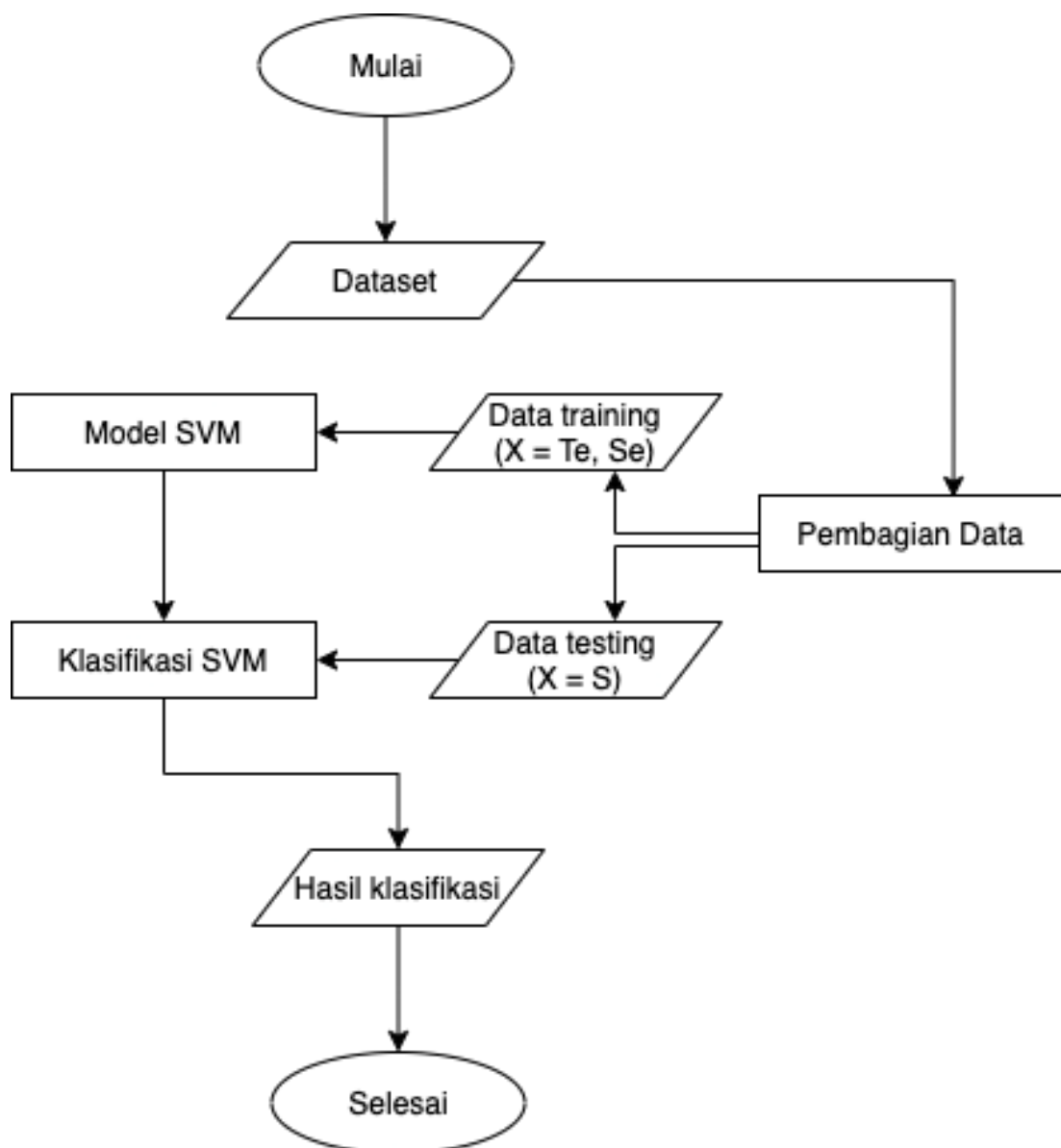
Gambar 2. Proses kalkulasi DFT

Hasil dari perhitungan DFT menggunakan VASP menghasilkan beberapa berkas namun, informasi yang diperlukan untuk membuat fitur-fitur pada dataset terdapat pada berkas OUTCAR yang berisi informasi posisi-posisi atom yang sudah dioptimasi, momen magnetik, jumlah elektron dalam sistem, energi fermi, energi band, serta total energi.

Lalu untuk analisa dan validasi struktur, kami menggunakan program *p4vasp* untuk mengamati karakteristik suatu struktur hasil kalkulasi VASP, informasi ini bisa didapat melalui berkas OUTCAR, DOSCAR, CHGCAR, serta *vasprun.xml*.

3.3. Gambaran Umum Sistem

Berikut adalah gambaran umum dari sistem pencari bahan magnet kuat yang akan dibangun. Langkah pertama yang akan dilakukan adalah membangun dataset yang dibutuhkan yaitu 132 struktur $X = \text{Te, Se}$ dan 66 struktur $X = \text{S}$ menggunakan program VASP, setiap kalkulasi DFT yang dilakukan bertujuan untuk mendapatkan karakteristik bahan magnet yang konvergen dari setiap kombinasi struktur, data hasil kalkulasi DFT ini kemudian akan dibuat dataset yang kemudian akan digunakan untuk membangun model klasifikasi SVM. Model klasifikasi SVM akan digunakan untuk memprediksi label pada data uji ($X = \text{S}$). Sistem pencarian bahan magnet kuat yang akan dibangun menerima input berupa momen magnetik setiap atom (Tabel 3), posisi atom (Tabel 4), jumlah elektron dalam sistem, dan total energi dari suatu konfigurasi bahan dari struktur $X = \text{S}$ dan mengeluarkan output berupa label hasil klasifikasi.



Gambar 3. Gambaran umum sistem

3.4. Pembelajaran Mesin

Gambar 3 menjelaskan gambaran umum proses pembangunan sebuah model klasifikasi menggunakan metode *support vector machine* (SVM). Untuk menghindari masalah *overfitting*, model yang dibangun akan divalidasi menggunakan teknik *K-Fold Cross Validation* (CV). Teknik *K-Fold CV* memastikan bahwa evaluasi model yang dihasilkan tidak terlalu cocok (*overfit*) untuk hanya satu kelas data sehingga pemilihan model dapat lebih valid. Dengan menggunakan validasi *K-Fold CV*, model dipilih dan dievaluasi melalui *k* kali iterasi. *K-Fold CV* membagi dataset menjadi sejumlah *k* bagian atau lipatan di mana setiap lipatan digunakan sebagai set pengujian di beberapa titik. Pada proses klasifikasi, struktur $X = S$ akan diklasifikasikan, struktur $X = S$ memiliki fitur momen magnetik setiap atom (Tabel 3), posisi atom (Tabel 4), jumlah elektron dalam sistem, energi fermi, energi band, serta total energi kemudian sistem akan mengeluarkan label hasil klasifikasi yang didapatkan untuk konfigurasi bahan tersebut.

Sebelum melakukan pembelajaran menggunakan SVM, perlu dilakukan *scaling* serta normalisasi data ke dalam rentang nilai [0..1]. Hal ini perlu dilakukan karena SVM menggunakan nilai-nilai fitur yang berupa bilangan *float* untuk mengklasifikasikan data serta dapat menghindari apabila terdapat sampel baru yang memiliki rentang nilai yang sangat jauh yang dapat mempengaruhi kalkulasi SVM.

Model klasifikasi SVM yang dibangun akan mencari nilai-nilai parameter paling optimal dengan menggunakan teknik *Grid Search* berbasis *K-Fold Cross Validation* dengan nilai $K=10$. Dengan teknik *Grid Search*, model SVM yang dibangun melakukan proses pencarian parameter optimal dari setiap nilai yang diperlukan setiap *kernel* (Tabel 4). Untuk dapat melakukan proses pencarian, *Grid Search* memerlukan rentang nilai untuk setiap parameter. Rentang nilai-nilai parameter yang digunakan pada penelitian ini antara lain nilai $C = [0.1, 1, 10, 100, 1000]$, nilai $\gamma = [0.1, 0.01, 0.001, 0.0001, 0.00001]$, nilai $d = [1, 2, 3, 4, 5]$, serta dengan tiga kernel yang berbeda yaitu *Linear*, *RBF*, dan *Polynomial*. Perlu diketahui bahwa teknik *Grid Search* bersifat *exhaustive search* dengan kata lain akan mencoba semua kemungkinan kombinasi parameter yang ada, jika jumlah kombinasi yang diberikan sangat banyak, maka, hal ini dapat sangat berdampak pada waktu proses pencarian parameter terbaik. Proses ini dinamakan *hyperparameter tuning*.

Seleksi fitur berdasarkan nilai korelasi dengan kelas juga akan diterapkan untuk observasi lebih lanjut mengenai fitur-fitur yang paling berpengaruh terhadap penentuan kelas. Nilai-nilai korelasi ini didapat dengan menggunakan *Pearson correlation coefficient*. Dengan demikian, akan terdapat beberapa skenario pengujian kinerja model klasifikasi SVM, tentu saja dengan menggunakan *hyperparameter* yang berbeda untuk setiap set fitur.

3.4. Evaluasi

Model klasifikasi yang dibangun akan dievaluasi untuk diukur kualitasnya dengan menggunakan *confusion matrix*. Salah satu langkah paling umum dan intuitif yang bisa dilakukan adalah menghitung akurasi dengan membandingkan label set data uji aktual dengan label hasil klasifikasi, sehingga dapat diformulasikan dengan:

$$A = \frac{(P + N)}{(P + N + P + N)} \tag{9}$$

seiring dengan akurasi, beberapa metrik lain yang dapat digunakan untuk mengevaluasi kinerja model klasifikasi yang akan digunakan adalah sebagai berikut:

1. *Precision* akan mengukur ketepatan identifikasi dari sistem dalam mengklasifikasikan data *true positive* dari total data yang diprediksi positif. Dihitung sebagai jumlah prediksi positif yang benar dibagi dengan jumlah total sampel yang diprediksi positif.

$$P = \frac{P}{P + \Phi P} \tag{10}$$

2. *Recall* merupakan ketepatan identifikasi dari sistem dalam mengklasifikasikan data *true positive*, dihitung sebagai jumlah prediksi positif yang benar dibagi dengan jumlah total sampel positif.

$$R = \frac{P}{P + N} \tag{11}$$

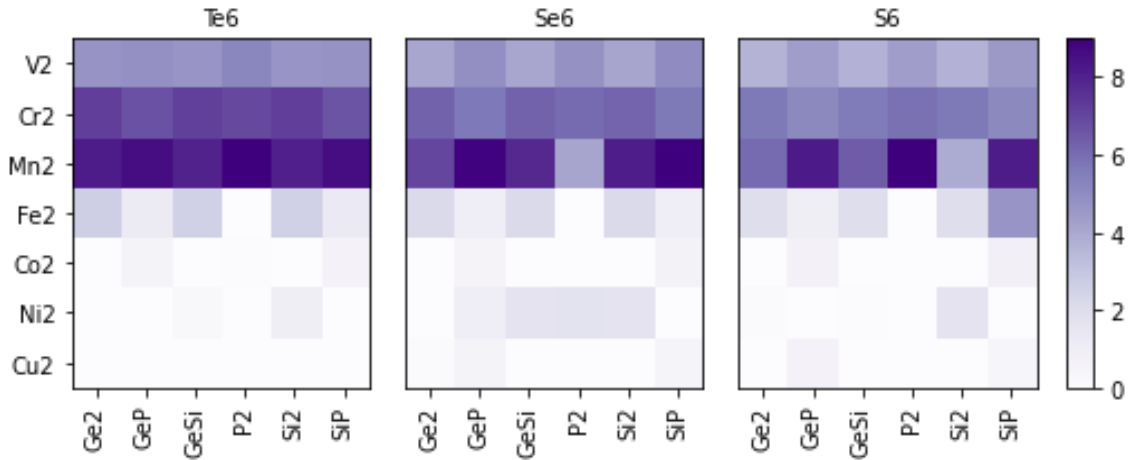
3. *F1-Score* merupakan nilai rata-rata dari *recall* dan *precision*, yang diberi bobot.

$$\begin{matrix} \diamond & \diamond & \diamond & \diamond & \diamond & \diamond \\ \diamond & \diamond & \diamond & \diamond & \diamond & \diamond \\ \diamond & \diamond & \diamond & \diamond & \diamond & \diamond \\ \diamond & \diamond & \diamond & \diamond & \diamond & \diamond \end{matrix} = \begin{matrix} \diamond & * & \frac{P \diamond \diamond \diamond \diamond \diamond \diamond}{P \diamond \diamond \diamond \diamond \diamond \diamond + \diamond \diamond \diamond \diamond} \end{matrix} \quad (12)$$

4. Hasil dan Analisis

4.1. Hasil Simulasi *Density Functional Theory*

Berikut adalah sebaran momen magnetik hasil simulasi *Density Functional Theory* (DFT) dari setiap kombinasi struktur $A_2B_2X_6$.

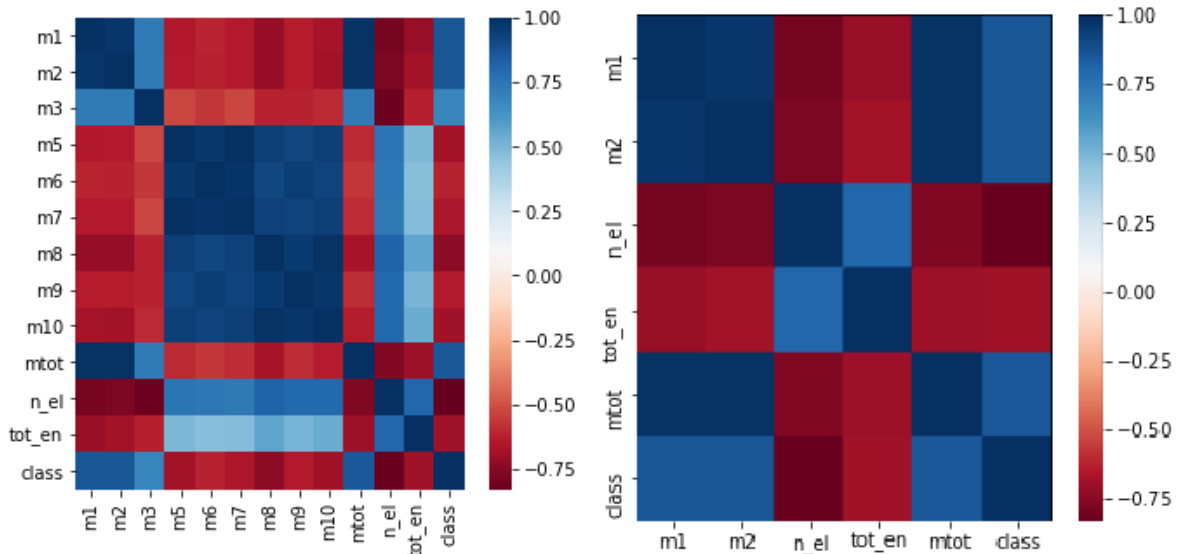


Gambar 4. Momen magnetik (dalam μB)

Sifat magnetik struktur seperti momen magnetik, sangat dipengaruhi oleh momen magnetik atom-atom penyusun pada situs A. Struktur yang dibuat dengan mensubtitusi atom yang memiliki momen magnetik besar pada situs A menghasilkan serangkaian struktur dengan variasi yang relatif besar dalam momen magnetiknya. Sebaliknya pada atom di situs B dan X, atom-atom di situs ini memiliki kontribusi yang dapat diabaikan terhadap momen magnetik keseluruhan pada suatu struktur.

4.2. Hasil Klasifikasi *Support Vector Machine*

Pada penelitian ini, kinerja model klasifikasi yang dibangun menggunakan *support vector machine* (SVM) akan di evaluasi dengan menggunakan tiga kernel yang berbeda yaitu kernel *Linear*, *Radial Basis Function* (RBF) atau *Gaussian*, dan *Polynomial*. Pengujian akan dilakukan dengan menggunakan dataset dengan fitur lengkap serta dengan menyeleksi fitur yang memiliki korelasi cukup tinggi terhadap kelas. Seleksi fitur di sini dilakukan dengan menggunakan matriks korelasi (Gambar 5). Fitur yang diseleksi adalah fitur yang memiliki nilai korelasi sekitar 0.5 serta -0.5, dengan kata lain, fitur yang memiliki nilai korelasi mendekati 0 mengartikan bahwa fitur tersebut memiliki korelasi yang lemah atau bahkan tidak memiliki korelasi sama sekali dengan penentuan kelas.



Gambar 5. Korelasi fitur

Dengan menerapkan seleksi fitur, kami dapat memilih 12 dari total 45 fitur yang memiliki nilai korelasi cukup tinggi terhadap kelas yaitu momen magnetik pada atom-atom di situs A ["m1", "m2"], atom pertama di situs B ["m3"], semua atom pada situs X ["m5", ..., "m10"], momen magnetik total dari suatu struktur ["mtot"], jumlah elektron pada sistem ["n_el"], serta total energi ["tot_en"]. Fitur yang telah diseleksi ini kemudian akan digunakan untuk membuat total tiga skenario yang berbeda. Pertama kami akan menggunakan semua fitur yang ada, kedua, kami akan menggunakan fitur yang telah kami seleksi yang telah disebutkan sebelumnya, lalu yang terakhir, kami akan menggunakan fitur yang memiliki korelasi positif dan negatif tertinggi terhadap kelas. Dari ketiga skenario yang telah disebutkan, kami akan melakukan observasi terhadap pengaruh seleksi fitur dengan kinerja model SVM yang dibangun. *Hyperparameter* serta kinerja model SVM yang dibangun masing-masing akan ditunjukkan pada Tabel 5 dan Tabel 6 beserta dengan nilai akurasi, *precision*, *recall* serta *f1-Score* yang didapat dari setiap skenario.

Tabel 5. Tabel parameter optimal tiap skenario

Jumlah Fitur	Kernel	Hyperparameter			Grid Score
		C	γ	d	
45	Polynomial	1000	-	3	91.66%
12	Linear	0.1	-	-	97.61%
5	RBF	0.1	0.1	-	92.85 %

Tabel 6. Tabel skenario pengujian

Jumlah Fitur	TP	FP	TN	FN	Akurasi	Precision	Recall	F1-Score
45	21	0	19	2	95.23%	100%	91.30%	95.45%
12	21	0	19	2	95.23%	100%	91.30%	95.45%
5	22	0	19	1	97.61%	100%	92.56%	97.77%

4.3. Analisis Hasil

Berdasarkan hasil pengujian menggunakan model SVM yang dibangun (Tabel 6), dapat diamati bahwa momen magnetik pada atom-atom pada situs A ["m1", "m2"], atom pertama pada situs B ["m3"], semua atom pada situs X ["m5", ..., "m10"], momen magnetik total dari suatu struktur ["mtot"], jumlah elektron pada sistem ["n_el"], serta total energi ["tot_en"] yang telah diseleksi dapat menghasilkan kinerja yang sama dengan menggunakan semua fitur. Seleksi terhadap beberapa fitur yang memiliki nilai korelasi positif dan negatif paling tinggi seperti atom-atom pada situs A ["m1", "m2"], jumlah elektron pada sistem ["n_el"], serta total energi ["tot_en"] dapat meningkatkan kinerja model SVM dalam mengklasifikasikan magnet kuat dengan meningkatnya nilai *recall* serta jumlah *true positive* diiringi dengan turunnya jumlah *false negative*.

Hal ini menunjukkan bahwa pada dataset yang telah dibuat, bahan magnet dapat diklasifikasikan secara akurat dengan mengamati nilai momen magnetik atom pada situs A, momen magnetik total dari struktur, jumlah elektron pada sistem, serta total energi. Berdasarkan dari semua skenario pengujian yang telah dilakukan (Tabel 6), dapat ditunjukkan bahwa model klasifikasi SVM yang dibangun mampu mendapatkan nilai akurasi serta *f1-score* paling tinggi masing-masing sebesar 97.61% dan 97.77% menggunakan kernel RBF, dengan menyeleksi fitur-fitur yang memiliki nilai korelasi paling tinggi terhadap kelas, serta nilai $C = 0.1$ dan $\gamma = 0.1$.

5. Kesimpulan

Dari hasil pengujian sistem pencarian bahan magnet ini, didapatkan beberapa kesimpulan sebagai berikut:

1. Metode alternatif dengan menggunakan SVM terbukti dapat mengakselerasi proses pencarian bahan magnet kuat melalui klasifikasi bila dibandingkan dengan simulasi DFT (*Hard Computing*).
2. Model klasifikasi SVM yang dibangun dapat secara akurat mengklasifikasikan magnet kuat dengan menerapkan seleksi fitur yang tepat
3. Model pembelajaran mesin yang dibangun dapat digunakan untuk mengklasifikasikan magnet kuat berdasarkan nilai akurasi serta *f1-score* terbaik yang cukup tinggi yaitu sekitar 97%

Daftar Pustaka

- [1] T. D. Rhone, W. Chen, S. Desai, A. Yacoby and E. Kaxiras, "Data-driven studies of magnetic two-dimensional materials," 2018.
- [2] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder and K. A. Persson, "Commentary: The Materials Project: A materials genome approach to accelerating materials innovation," *APL Materials*, vol. 1, 2013.
- [3] S. A. Tawfik, O. Isayev, C. Stampfl, J. Shapter, D. A. Winkler and M. J. Ford , "Efficient Prediction of Structural and Electronic Properties of Hybrid 2D Materials Using Complementary DFT and Machine Learning Approaches," *Advanced Theory and Simulations*, vol. 2, no. 1, 2018.
- [4] R. Schleder, Gabriel and Padilha, Antonio Claudio and Acosta, Carlos and Costa, Marcio and Fazzio and Adalberto, "From DFT to Machine Learning: recent approaches to Materials Science –a review," *Journal of Physics: Materials*, 2019.
- [5] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, p. 273–297, 1995.
- [6] P. Hohenberg and W. Kohn, "Inhomogeneous Electron Gas," *Phys. Rev.*, vol. 136, no. 3B, pp. B864-B871, 1964.
- [7] W. Kohn and L. J. Sham, "Self-Consistent Equations Including Exchange and Correlation Effects," *Phys. Rev.*, vol. 140, no. 4A, pp. A1133-A1138, 1965.
- [8] A. Seko, H. Hayashi, K. Nakayama, A. Takahashi and I. Tanaka, "Representation of compounds for machine-learning prediction of physical properties," *Phys. Rev.*, vol. 95, no. 14, p. 144110, 2017.
- [9] T. Nozaki, A. Koziol-Rachwał, M. Tsujikawa, Y. Shiota, X. Xu, T. Ohkubo, T. Tsukahara, S. Miwa, M. Suzuki, S. Tamaru, H. Kubota, A. Fukushima, K. Hono, M. Shirai and Y. Suzuki, "Highly efficient voltage control of spin and enhanced interfacial perpendicular magnetic anisotropy in iridium-doped Fe/MgO magnetic tunnel junctions," *NPG Asia Materials*, vol. 9, no. 12, pp. e451-e451, 2017.
- [10] W.-C. Lu, X.-B. Ji, M.-J. Li, L. Liu, B.-H. Yue and L.-M. Zhang, "Using support vector machine for materials design," *Advances in Manufacturing*, vol. 1, no. 2, p. 151–159, 2013.
- [11] G. Menichetti, M. Calandra and M. Polini, "Electronic structure and magnetic properties of few-layer Cr₂Ge₂Te₆: the key role of nonlocal electron–electron interaction effect," *2D Materials*, vol. 6, 2019.
- [12] Himanen, Lauri and Geurts, Amber and Foster, Adam Stuart and Rinke and Patrick, "Data-Driven Materials Science: Status, Challenges, and Perspectives," *Advanced Science*, vol. 6, no. 21, 2019.