



ITS
Institut
Teknologi
Sepuluh Nopember

TUGAS AKHIR - TM091585

ANALISA PENGARUH KOMPOSISI KIMIA TERHADAP SIFAT-SIFAT PENYIMPANAN HIDROGEN LOGAM PADUAN AB_2 MENGGUNAKAN METODE *MACHINE LEARNING*

GHAZY DICKY
NRP 0211 16 40000 013

Dosen Pembimbing
Suwarno ST., Msc., PhD.

JURUSAN TEKNIK MESIN
Fakultas Sistem dan Teknologi Industri
Institut Teknologi Sepuluh Nopember
Surabaya 2020



TUGAS AKHIR - TM091585

**ANALISA PENGARUH KOMPOSISI KIMIA TERHADAP
SIFAT-SIFAT PENYIMPANAN HIDROGEN LOGAM
PADUAN AB_2 MENGGUNAKAN METODE *MACHINE
LEARNING***

GHAZY DICKY
NRP 0211 16 40000 013

Dosen Pembimbing
Suwarno ST., Msc., PhD.

JURUSAN TEKNIK MESIN
Fakultas Sistem dan Teknologi Industri
Institut Teknologi Sepuluh Nopember
Surabaya 2020



FINAL ASSIGNMENT - TM184730

MACHINE LEARNING ANALYSIS OF THE ALLOYING ELEMENT EFFECTS ON AB₂ METAL HYDRIDES HYDROGEN STORAGE PROPERTIES

GHAZY DICKY
NRP 0211 16 40000 013

Supervisor
Suwarno ST., Msc., PhD.

MECHANICAL ENGINEERING DEPARTMENT
Faculty of Industrial System and Technology
Institut Teknologi Sepuluh Nopember
Surabaya 2020

LEMBAR PENGESAHAN
ANALISA PENGARUH KOMPOSISI KIMIA TERHADAP
SIFAT-SIFAT PENYIMPANAN HIDROGEN LOGAM
PADUAN AB₂ MENGGUNAKAN METODE MACHINE
LEARNING

TUGAS AKHIR

Diajukan Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Teknik
pada
Bidang Studi Metalurgi
Program Studi S-1 Departemen Teknik Mesin
Fakultas Teknologi Industri dan Rekayasa Sistem
Institut Teknologi Sepuluh Nopember

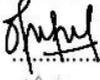
Oleh:



Ghazy Dicky

NRP. 02111640000013

Disetujui oleh Pembimbing Tugas Akhir:

1. Suwarno, S.T., M.Sc., Ph.D.
NIP. 198005202005011003  (Pembimbing)
2. Fahmi Mubarok, S.T., M.Sc., Ph.D.
NIP. 197801152003121002  (Penguji I)
3. Mohammad Khoirul Effendi, S.T., M.Sc. Eng.
NIP. 198204142010121001  (Penguji II)
4. Putu Suwarta, S.T., M.Sc., Ph.D.
NIP. 198203032009121007  (Penguji III)

SURABAYA
AGUSTUS, 2020

Halaman ini sengaja dikosongkan

**ANALISA PENGARUH KOMPOSISI KIMIA TERHADAP
SIFAT-SIFAT PENYIMPANAN HIDROGEN LOGAM
PADUAN AB₂ MENGGUNAKAN METODE *MACHINE
LEARNING***

Nama Mahasiswa : Ghazy Dicky
NRP : 0211164000013
Jurusan : Teknik Mesin
Dosen Pembimbing : Suwarno, ST., M.Sc., PhD.

Abstrak

Aplikasi hydrogen sebagai bahan bakar kendaraan bermotor membutuhkan hydrogen untuk disimpan pada tangki penyimpanan eksternal, dimana logam hidrida adalah material yang memiliki potensi tinggi untuk menyimpan hydrogen. Dataset logam hidrida AB₂ yang terdiri dari komposisi kimia dari logam paduan beserta sifat-sifat penyimpanan hydrogen yang didapatkan dari 50 jurnal penelitian terdahulu dianalisa menggunakan metode machine learning. Prediksi pengaruh dari unsur paduan terhadap energi panas pembentukan (ΔH dan ΔS), phase abundance, dan kapasitas hydrogen (wt%), serta klasifikasi pengaruh dari setiap unsur serta nilai B/A terhadap sifat penyimpanan hidrogen didapatkan dari analisa yang dilakukan menggunakan tiga algoritma machine learning, yaitu multivariate linear regression, decision tree, dan random forest. Skor R^2 dan metric error diperoleh dari evaluasi model yang dilakukan. Model random forest menghasilkan skor R^2 tertinggi dan error terendah untuk memprediksi seluruh sifat penyimpanan hydrogen pada dataset dengan unsur-unsur independen. Skor R^2 tertinggi dan error terendah diraih model decision tree untuk prediksi delta entropi, phase abundance, dan persen massa hydrogen, serta model random forest untuk prediksi delta entalpi pada dataset dengan unsur-unsur yang dikelompokkan menjadi unsur A dan B.

Kata kunci: logam paduan AB₂, sifat-sifat hidrogenasi, machine learning.

Halaman ini sengaja dikosongkan

MACHINE LEARNING ANALYSIS OF THE ALLOYING ELEMENT EFFECTS ON AB₂ METAL HYDRIDES HYDROGEN STORAGE PROPERTIES

Name : Ghazy Dicky
NRP : 0211164000013
Department : Mechanical Engineering
Supervisor : Suwarno, ST., M.Sc., PhD.

Abstract

Hydrogen is one of the most potential clean and sustainable energy sources to be utilized. The application of hydrogen as fuels requires hydrogen to be stored in several forms, with metal hydrides being the form of storage with high potency. The alloying elements of metal hydrides are frequently studied to discover alloy with the most beneficial and desired properties. However, until recently, standard methods to ascertain the alloying composition is yet present, leaving researchers with the logical design based alloying element determination. In an effort to assist researchers, machine learning approaches are developed to analyze metal hydrides dataset consisting of chemical compositions of alloys as well as hydrogen storage properties obtained from 50 previous research journals, aiming to predict the effect of the alloying element on the heat of formation (ΔH and ΔS), phase abundance, and the hydrogen capacity (wt% H) of the alloy as well as to reveal the importance of the features. Three models are being employed namely multivariate linear regression, decision tree, and random forest which were evaluated by comparing predicted values with the actual values resulting in R^2 scores and error metrics. The random forest model generates the best score out of all machine learning models.

Keywords: Machine Learning, Metal Hydrides, Hydrogen Storage, Hydrogen Energy, AB₂ Alloys.

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Puji syukur kepada Allah SWT berkat Rahmat, Hidayah, dan Karunia-Nya kepada kita semua sehingga penulis dapat menyelesaikan tugas akhir dengan judul “**Analisa Pengaruh Komposisi Kimia terhadap Sifat-Sifat Penyimpanan Hidrogen Logam Paduan AB₂ Menggunakan Metode *Machine Learning***”.

Pada kesempatan ini penulis ingin mengucapkan terima kasih kepada:

1. Ayah saya **Yackie**, bunda saya **Dewi Safitri**, dan adik saya **Nadya Yasmin Dicky** atas restu, pengorbanan dan doanya hingga saya dapat menempuh pendidikan hingga tahap sarjana.
2. Bapak **Suwarno, ST., M.Sc., Ph.D.**, selaku Dosen Pembimbing atas bimbingan, saran, dan motivasi yang diberikan.
3. Bapak **Fahmi Mubarak, ST., M.Sc., Ph.D.**, **Putu Suwarta, S.T., M.Sc., Ph.D.**, **Mohammad Khoirul Effendi, S.T., M.Sc.Eng., Ph.D.**, dan **Abdillah Suyuthi, Ph.D.**, atas bimbingan, saran, dan motivasi yang diberikan.
4. **Teman-teman seperjuangan tugas akhir laboratorium metalurgi** atas dukungan dan bantuannya.
5. Teman-teman **M-59** atas kebersamaan dan persaudaraannya hingga saat ini.
6. **Sufjan Stevens, Ben Platt, Dua Lipa**, dan **George Ezra**, atas hiburan dan motivasinya hingga buku ini dapat diselesaikan.
7. Seluruh **civitas akademika Jurusan Teknik Mesin FTI-ITS** yang telah memberikan dukungan moril kepada penulis.

Penulis menyadari tugas akhir ini tidak luput dari berbagai kekurangan. Penulis mengharapkan saran dan kritik demi kesempurnaan dan perbaikannya sehingga akhirnya laporan tugas akhir ini dapat memberikan manfaat bagi bidang pendidikan dan penerapan dilapangan serta dapat dikembangkan lagi lebih lanjut.

Halaman ini sengaja dikosongkan

DAFTAR ISI

LEMBAR PENGESAHAN	i
ABSTRAK	iii
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xv
BAB I PENDAHULUAN	1
1.1. Latar belakang	1
1.2. Rumusan Masalah	3
1.4. Tujuan.....	3
1.5. Manfaat Program.....	4
1.6. Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA	5
2.1. Hidrogen Sebagai Sumber Energi Kendaraan Bermotor....	5
2.2. Teknologi Penyimpanan Hidrogen	6
2.3. Material Penyimpan Hidrogen	9
2.3.1. Hydrogen container berbasis logam hidrida.....	10
2.3.2 Logam Hidrida (Metal Hydride).....	11
2.4. Hidrogenasi Logam Hidrida	12
2.5. Logam Hidrida AB_2	14

2.6. Penerapan Metode <i>Machine Learning</i> pada Ilmu Material	16
2.6.1. Prinsip dasar <i>Machine Learning</i>	16
2.6.2. <i>Database</i>	18
2.6.3. Fitur	18
2.6.4. Algoritma Multivariate Regression	19
2.6.7. Algoritma Decision Tree	20
2.6.8. Algoritma Random Forest	23
2.6.9. Interpretabilitas dari Metode <i>Machine Learning</i>	24
2.7. Penelitian Terdahulu	25
BAB III METODE PENELITIAN.....	29
3.1. Diagram Alir Penelitian.....	29
3.2. Data dan Sumber Data.....	30
3.3. Desain Eksperimen	30
3.4. Pemilihan Algoritma	32
3.5. Analisa Awal dan Visualisasi Data	32
BAB IV HASIL DAN PEMBAHASAN	39
4.1. Data Awal	39
4.2. Hasil Pengolahan Awal Data.....	41
4.2.1. <i>Feature Engineering</i>	41
4.1.2. Preprocessing Data	44
4.3. Modelling Variabel dengan Unsur-Unsur yang Tidak Dikelompokkan	50

4.3.1. Perbandingan Hasil Evaluasi Algoritma untuk Prediksi Selisih Entalpi.....	50
4.3.2. Perbandingan Hasil Evaluasi Algoritma untuk Prediksi Selisih Entropi	54
4.3.3. Perbandingan Hasil Evaluasi Algoritma untuk Prediksi Phase Abundance C14.....	56
4.3.4. Perbandingan Hasil Evaluasi Algoritma untuk Prediksi Persen Massa Hidrogen	58
4.3.5. Pengaruh Unsur terhadap Heat of Formation	61
4.3.6. Pengaruh Unsur terhadap Phase Abundance C14	66
4.3.7. Pengaruh Unsur terhadap Persen Massa Hidrogen....	69
4.4. Modelling Variabel dengan Unsur-Unsur yang Dikelompokkan menjadi A dan B	71
4.4.1. Hasil Multivariate Regression	71
4.4.2. Hasil Decision Tree	72
4.4.3. Hasil Random Forest	73
4.4.4. Pengaruh Rasio B/A terhadap Variabel Output.....	75
BAB V KESIMPULAN DAN SARAN.....	79
5.1. Kesimpulan.....	79
5.2. Saran.....	79
DAFTAR PUSTAKA	81
LAMPIRAN.....	85

DAFTAR TABEL

Tabel 2.1 Metode Penyimpanan Hidrogen.....	8
Tabel 2.2 Properti Utama Logam Hidrida untuk Aplikasi Fase Gas.....	12
Tabel 3.1 Koefisien Regresi Linear Prediksi Delta H.....	33
Tabel 3.2 Koefisien Regresi Linear Prediksi Delta S.....	34
Tabel 4.1 Nilai Kemampuan Prediksi Seluruh Variable Target untuk Masing-masing Variasi Fitur.....	36
Tabel 4.2 Perbandingan Metode Machine Learning untuk Prediksi Selisih Entalphi Pembentukan.....	49
Tabel 4.3 Perbandingan Metode Machine Learning untuk Prediksi Selisih Entropi Pembentukan.....	52
Tabel 4.4 Perbandingan Metode Machine Learning untuk Prediksi Phase Abundance C14.....	55
Tabel 4.5 Perbandingan Metode Machine Learning untuk Prediksi Persen Massa Hidrogen.....	57
Tabel 4.6 Terjemahan Decision Tree untuk Selisih Entalphi Absorpsi.....	61
Tabel 4.7 Terjemahan Decision Tree untuk Selisih Entropi Absorpsi.....	63
Tabel 4.8 Terjemahan Decision Tree untuk Phase Abundance...	65
Tabel 4.9 Terjemahan Decision Tree untuk Persen Massa Hidrogen.....	68

Tabel 4.10 Hasil Regresi Multivariate untuk unsur A dan B yang dikelompokkan.....	69
Tabel 4.11 Hasil Regresi Decision Tree untuk unsur A dan B yang dikelompokkan.....	70
Tabel 4.12 Hasil Regresi Random Forest untuk unsur A dan B yang dikelompokkan.....	71
Tabel A.1 Hasil Regresi Linear Multivariate untuk Setiap Variabel Output.....	82
Tabel A.2 Hasil Regresi Decision Tree untuk Setiap Variabel Output.....	82
Tabel A.3 Hasil Regresi Random Forest untuk Setiap Variabel Output.....	82
Tabel B.1 Hasil Regresi Pengaruh Komposisi Kimia Terhadap Selisih Entalpi Absorpsi.....	83
Tabel B.2 Ranking Pengaruh Komposisi Kimia terhadap Selisih Entalpi Absorpsi berdasarkan Regresi Linear.....	83
Tabel B.3 Hasil Regresi Pengaruh Komposisi Kimia Terhadap Selisih Entropi Absorpsi.....	84
Tabel B.4 Ranking Pengaruh Komposisi Kimia terhadap Selisih Entropi Absorpsi berdasarkan Regresi Linear.....	84
Tabel B.5 Hasil Regresi Pengaruh Komposisi Kimia Terhadap Phase Abundance C14.....	85
Tabel B.6 Ranking Pengaruh Komposisi Kimia terhadap Phase Abundance C14 berdasarkan Regresi Linear.....	86

Tabel B.7 Hasil Regresi Pengaruh Komposisi Kimia Terhadap
Persen Massa Hidrogen..... 86

Tabel B.8 Ranking Pengaruh Komposisi Kimia terhadap Persen
Massa Hidrogen berdasarkan Regresi Linear..... 87

DAFTAR GAMBAR

Gambar 2.1 Sel Bahan Bakar Hidrogen.....	5
Gambar 2.2 Perbandingan kontainer hidrogen dari segi ukuran fisik.....	10
Gambar 2.3 Pressure Composition Isotherm.....	13
Gambar 2.4 Workflow Supervised Learning.....	16
Gambar 2.5 Penentuan Kestabilan Atom menggunakan Decision Tree.....	20
Gambar 3.1 Diagram Alir Penelitian.....	28
Gambar 3.2 Ilustrasi Penelitian Secara Garis Besar	29
Gambar 3.3 Grafik Visualisasi Prevalensi Elemen pada Dataset.....	32
Gambar 3.4 Plot Variabel Input dengan Variabel Output.....	35
Gambar 4.1 Prevalensi Elemen pada Dataset Awal.....	38
Gambar 4.2 Prevalensi Elemen pada Dataset untuk 8 Unsur dengan Prevalensi Terbesar.....	39
Gambar 4.3 Contoh Dataset untuk Variabel Input dengan Unsur-Unsur yang Tidak Dikelompokkan.....	40
Gambar 4.4 Contoh Dataset dengan Unsur-Unsur yang Dikelompokkan Menjadi Unsur A dan B.....	42
Gambar 4.5 Pengecekan <i>Null</i>	44

Gambar 4.6 Heat Map Uji Kolerasi Variabel Input.....	44
Gambar 4.7 Deteksi dan Penanganan Outliers.....	46
Gambar 4.8 Jumlah Data Set untuk Setiap Variabel.....	46
Gambar 4.9 <i>Data Scaling</i>	47
Gambar 4.10 Perbandingan Data Sebelum Scaling dan Sesudah Scaling.....	47
Gambar 4.11 Grafik Nilai Asli dan Nilai Prediksi untuk Selisih Entalphi Pembentukan.....	50
Gambar 4.12 Grafik Nilai Asli dan Nilai Prediksi untuk Selisih Entropi Pembentukan	53
Gambar 4.13 Grafik Nilai Asli dan Nilai Prediksi untuk Phase Abundance C14.....	55
Gambar 4.14 Grafik Nilai Asli dan Nilai Prediksi untuk Persen Massa Hidrogen.....	58
Gambar 4.15 Grafik Pengaruh Elemen untuk ΔH	60
Gambar 4.16 Grafik Pengaruh Elemen untuk ΔS	62
Gambar 4.17 Grafik Pengaruh Elemen untuk Phase Abundance..	64
Gambar 4.18 Grafik Pengaruh Elemen untuk Kapasitas Hidrogen.....	67
Gambar 4.19 Pengaruh rasio B/A terhadap variable.....	73
Gambar 4.20 Pengaruh rasio B/A terhadap variable delta entropi.	73

Gambar 4.21 Pengaruh rasio B/A terhadap variable phase abundance C14.....	74
Gambar 4.22 Pengaruh rasio B/A terhadap variable persen massa hidrogen	74
Gambar B.2 Klasifikasi Pengaruh Komposisi Kimia terhadap Selisih Entalphi Absorpsi berdasarkan Decision Tree.....	84
Gambar B.3 Klasifikasi Pengaruh Komposisi Kimia terhadap Selisih Entropi Absorpsi berdasarkan Decision Tree.....	85
Gambar B.4 Klasifikasi Pengaruh Komposisi Kimia terhadap Phase Abundance C14 berdasarkan Decision Tree.....	86
Gambar B.5 Klasifikasi Pengaruh Komposisi Kimia terhadap Persen Massa Hidrogen berdasarkan Decision Tree.....	87
Gambar C.1 Pengecekan Jumlah Tree Optimum untuk Variabel Selisih Entalphi Pembentukan.....	88
Gambar C.2 Pengecekan Jumlah Tree Optimum untuk Variabel Selisih Entropi Pembentukan.....	88
Gambar C.3 Pengecekan Jumlah Tree Optimum untuk Variabel Phase Abundance C14.....	89
Gambar C.4 Pengecekan Jumlah Tree Optimum untuk Variabel Persen Massa Hidrogen.....	89

Halaman ini sengaja dikosongkan

BAB I

PENDAHULUAN

1.1. Latar belakang

Transisi penggunaan energi menuju energi yang lebih bersih dan sustainable untuk mengatasi ketergantungan penggunaan bahan bakar fosil yang semakin langka mendasari banyaknya riset di bidang energi alternatif. Perlu dikembangkan suatu energi alternatif yang memiliki prospek menjanjikan dimana hidrogen merupakan salah satu energi alternatif yang saat ini banyak diteliti oleh kalangan ilmuwan karena dapat dikonversi menjadi energi listrik dengan sistem sel bahan bakar atau fuel cell. Produsen kendaraan bermotor telah mengembangkan mobil berbasis fuel cell, namun fuel cell memerlukan suatu tangki penyimpanan eksternal yang menjadi hambatan pengaplikasian fuel cell pada kendaraan bermotor. Tangki penyimpanan ini berfungsi untuk menampung gas hidrogen, sama halnya seperti tangki bahan bakar minyak pada mobil komersil.

Senyawa AB_2 merupakan kandidat potensial sebagai material penyerap hidrogen untuk aplikasi pada kendaraan yang menggunakan fuel cell karena secara teoritis memiliki sifat-sifat yang diinginkan untuk penerapan pada kendaraan bermotor, seperti kemampuan menyerap hidrogen dalam jumlah yang cukup tinggi (~3.8 wt%) dibandingkan dengan material penyimpan hidrogen lainnya seperti AB_5 , temperatur operasi dibawah 100 derajat celcius, ketahanan terhadap korosi, serta kinetik absorpsi/desorpsi hidrogen yang cukup tinggi.

Dalam beberapa tahun terakhir telah dilakukan berbagai riset terhadap berbagai paduan senyawa AB_2 guna mempelajari sifat-sifat dari paduan tersebut, baik dari sifat penyimpanan hidrogen seperti gravimetric density yang sangat penting bagi aplikasi *hydrogen storage* pada kendaraan bermotor karena tangki

tersebut harus bisa menyimpan hidrogen dalam kapasitas yang besar dengan ukuran tangki yang kecil serta ringan, maupun dari segi harga serta ketersediaan bahan baku seperti penelitian yang dilakukan oleh Ulmer et. al. [1] pada 2015 yang menunjukkan kemungkinan penurunan harga dari material AB_2 sebagai efek dari substitusi vanadium vanadium sebagai elemen A pada AB_2 dengan ferrovanadium yang berharga lebih murah, yang berhasil menekan harga material hingga 1/3 harga aslinya.

Tentu saja, terdapat berbagai variasi paduan yang dapat diteliti guna menemukan variasi paduan yang paling menguntungkan dimana setiap penggantian variasi paduan akan memberikan pengaruh terhadap sifat-sifat hidrogenasi dari material tersebut. Namun, hingga saat ini belum ada metode standar untuk merancang variasi dari material AB_2 . Kemungkinan variasi paduan dari segi kimia dikombinasikan dengan parameter operasi dari sistem membuat desain logis untuk menentukan variasi ini menjadi sangat kompleks. Hal ini merupakan salah satu problema utama bagi para peneliti ditambah biaya penelitian yang tidak sedikit. Oleh karena itu diperlukan suatu cara untuk memprediksi variasi yang dapat memberikan sifat yang paling menguntungkan bagi material penyimpan hidrogen.

Sebelumnya telah dilakukan beberapa penelitian menggunakan algoritma *machine learning* guna memanfaatkan data histori untuk melakukan berbagai macam prediksi yang cepat dan tepat. Metode ini juga turut digunakan di penelitian material yang mulai mengubah paradigma klasik dari proses perhitungan serta simulasi material karena terdapat banyak data andal yang dapat digunakan. Suwarno dan Shahab [2] pada tahun 2004 melakukan penelitian metode jaringan syaraf tiruan (JST) guna memprediksi derajat presipitasi karbida krom pada baja tahan karat austenitik, penelitian ini menunjukkan bahwa JST berhasil melakukan prediksi dari presipitasi tersebut. Rahnama et al. [3] pada tahun 2019 melakukan penelitian berbasis machine learning

guna memprediksi presentase massa hidrogen pada logam hidrida. Metoda yang sama dilakukan pada penelitian ini guna memprediksi properti dari senyawa AB_2 .

1.2. Rumusan Masalah

Adapun rumusan masalah dari penelitian ini adalah sebagai berikut:

1. Bagaimana prediksi pengaruh komposisi kimia terhadap sifat-sifat penyimpanan hidrogen logam paduan AB_2 dengan metode peramalan *multivariate regression*, *decision tree*, dan *random forest* untuk elemen-elemen yang tidak dikelompokkan serta elemen-elemen yang dikelompokkan menjadi elemen A dan elemen B?
2. Bagaimana pengaruh elemen secara individu serta rasio B/A terhadap sifat-sifat penyimpanan hydrogen?

1.3. Batasan Masalah

Adapun batasan masalah dari penelitian ini adalah:

1. *Machine Learning* menganalisa properti logam paduan AB_2 berdasarkan data penelitian terdahulu,
2. Data yang digunakan adalah komposisi kimia, ΔH dan ΔS saat hidrogenasi dan dehidrogenasi, komposisi fasa mikrostruktur, dan persen massa hydrogen.
3. Verifikasi hasil prediksi model hanya akan dilakukan terhadap data validasi dan data testing, bukan terhadap logam paduan hasil eksperimen yang baru.

1.4. Tujuan

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Memprediksi pengaruh komposisi kimia terhadap sifat-sifat penyimpanan hidrogen logam paduan AB_2 dengan metode peramalan *multivariate regression*, *decision tree*, dan *random forest*, untuk elemen-elemen yang tidak dikelompokkan serta elemen-elemen yang dikelompokkan menjadi elemen A dan elemen B.

2. Mengetahui pengaruh elemen secara individu serta rasio B/A terhadap sifat-sifat penyimpanan hydrogen.

1.5. Manfaat Program

Manfaat program ini adalah untuk meningkatkan pengetahuan mengenai pengaruh komposisi kimia terhadap sifat-sifat hidrogenasi logam paduan AB_2 dengan metode peramalan *multivariate regression*, *decision tree*, dan *random forest*.

1.6. Sistematika Penulisan

Sistematika penulisan laporan penelitian Tugas Akhir dibagi dalam beberapa bab sebagai berikut:

1. Bab I Pendahuluan

Bab ini berisi tentang latar belakang, perumusan masalah, batasan masalah, tujuan, manfaat, serta sistematika penulisan penelitian tugas akhir.

2. Bab II Tinjauan Pustaka

Bab ini meliputi teori penunjang berisi semua hal yang menunjang dalam melakukan analisa hasil tugas akhir.

3. Bab III Metodologi Penelitian

Bab ini berisi uraian – uraian urutan proses pengerjaan tugas akhir ini dari awal sampai akhir.

4. Bab IV Analisa Data dan Pembahasan

Pada bab ini akan dijelaskan analisa hasil-hasil yang diperoleh selama pengujian dilakukan, pembahasan mengenai hubungan antara hasil pengujian dengan teori yang telah ada, dan berbagai macam analisa penunjang lain yang diperlukan.

5. Bab V Penutup

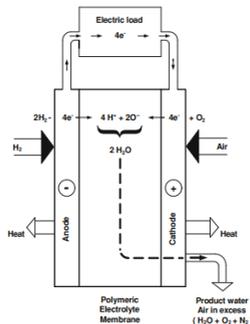
Bab penutup ini terdiri dari kesimpulan dan saran untuk pengembangan eksperimen dalam penelitian selanjutnya.

BAB II

TINJAUAN PUSTAKA

2.1. Hidrogen Sebagai Sumber Energi Kendaraan Bermotor

Ketergantungan terhadap konsumsi bahan bakar fosil pada kendaraan bermotor telah mengakibatkan menurunnya persediaan bahan bakar fosil di dunia serta menghasilkan polusi yang berdampak buruk bagi lingkungan. Hidrogen sebagai bahan bakar kendaraan bermotor merupakan salah satu sumber energi alternatif yang paling banyak diteliti oleh ilmuwan pada beberapa tahun silam. Hidrogen memiliki energi kimia per massa (142 MJ kg^{-1}) yang tiga kali lebih besar daripada bahan bakar kimia yang lain seperti bahan bakar minyak mentah (47 MJkg^{-1}). Selain itu hidrogen menawarkan proses sintesa bahan bakar yang ramah lingkungan dimana hidrogen akan menghasilkan uap air apabila dibakar dengan oksigen. Untuk menempuh jarak 400 km, sebuah mobil komersial modern mengkonsumsi sekitar 24 kg bahan bakar minyak di motor bakar, sedangkan untuk menempuh jarak yang sama mobil komersial modern hanya membutuhkan 8 kg hidrogen untuk versi motor bakar, dan 4 kg hidrogen untuk mobil listrik dengan sel bahan bakar (fuel cell) hidrogen [4]. Prinsip bekerjanya sel bahan bakar dapat dilihat pada ilustrasi dibawah ini.



Gambar 2.1 Sel Bahan Bakar Hidrogen

Pada ilustrasi diatas dapat dilihat bahwa hidrogen murni digunakan sebagai bahan bakar. Ion-ion hidrogen diproduksi pada reaksi oksidasi akan mengalir dari anoda menuju ke katoda melalui elektrolit, dimana elektron akan dipaksa untuk mengalir melalui sirkuit listrik eksternal agar dapat berguna menghasilkan arus listrik. Ion hidrogen akan diarahkan menuju elektrolit oleh perbedaan potensial yang dihasilkan dari oksidasi bahan bakar di anoda. Pada sisi katoda, elektron yang datang dari sirkuit eksternal akan bergabung dengan ion hidrogen dari sisi anoda dan oksigen dari udara untuk membentuk air. Uap air ini akan dibuang dari katode oleh aliran udara dan keluar bersama nitrogen beserta oksigen lebih yang tidak dipakau dalam reaksi elektrokimia secara keseluruhan. Hidrogen yang dipakai sebagai bahan bakar di sel bahan bakar ini disuplai dari tangki hidrogen eksternal.

Walau mobil listrik telah banyak dikembangkan, kendala yang dihadapi dalam aplikasi hidrogen sebagai bahan bakar mobil listrik adalah tangki penyimpanan eksternal hidrogen tersebut. Pada suhu ruangan dan tekanan atmosfer, 4 kg hidrogen akan menempati volume ruangan sebesar 45 m³ atau setara dengan sebuah tangki berbentuk bola dengan diameter 5m. Hal tersebut sangatlah tidak praktis untuk penggunaan di kendaraan bermotor. Oleh karena itulah dilakukan penelitian untuk menemukan material yang dapat menampung hidrogen dengan kapasitas yang tinggi dalam volume yang kecil serta dengan harga yang terjangkau.

2.2. Teknologi Penyimpanan Hidrogen

Prinsip dasar dari penyimpanan hidrogen adalah mereduksi volume gas hidrogen dalam ukuran besar untuk dimampatkan pada suatu tangki penyimpanan [5]. Hal yang perlu diperhatikan dari tangki penyimpanan hidrogen adalah material yang memiliki interaksi kuat dengan hidrogen dan tidak bereaksi apabila kontak dengan hidrogen, reversibility dimana kinetik harus tinggi pada temperatur yang rendah karena pada umumnya hidrogen hanya dapat dilepaskan apabila suhu berada diatas 800⁰

C atau apabila carbon dioksidasi. Tujuan utama dari teknologi penyimpanan hidrogen ini adalah untuk mengemas partikel hidrogen sedekat mungkin satu sama lain guna meraih kepadatan volumetrik yang tinggi dengan menggunakan material tambahan sesedikit mungkin. Sebuah sistem penyimpan hidrogen untuk aplikasi pada automotif setidaknya harus memenuhi beberapa persyaratan berikut ini, dengan memerhatikan sisi ekonomis dan lingkungan [6]:

1. kemampuan menyerap yang tinggi, baik secara gravimetric (>4.5 wt%) dan volumetric ($> 36\text{g H}_2/\text{L}$),
2. temperatur operasional yang moderat dalam rentang antara $60\text{--}120^\circ$ Celcius
3. reversibilitas dari siklus termal absorption/desorption, harga rendah,
4. tidak mudah tercemar (*low-toxicity*), dan
5. segi keamanan yang tinggi

Tangki bertekanan tinggi klasik yang biasanya dibuat dari baja murah memiliki batas tekanan 300 bar dan biasanya diisi hingga 200 bar di hampir semua negara. Untuk menyimpan 4 kg hidrogen diperlukan volume internal sebesar 225 liter atau 5 buah tangki dengan kapasitas 45 liter. Tangki bertekanan tinggi ini ketika penuh akan mengandung 4% hidrogen dari massa keseluruhan, namun dengan kerugian yang cukup signifikan, yaitu bahan bakar yang tersedia akan mengalami penurunan tekanan dari 450 bar menuju nol, sehingga diperlukan kontrol tekanan tambahan. Tangki bertekanan tinggi ini memiliki resiko yang cukup tinggi dimana kompresi bertekanan sangat tinggi ini adalah bagian yang paling berbahaya dan rumit, sehingga beberapa negara seperti Jepang melarang penggunaan tangki bertekanan tinggi pada mobil-mobil komersil. Kondensasi hidrogen menjadi liquid maupun hidrogen solid memberikan pandangan yang lebih menarik untuk meningkatkan massa per volume tangki. Massa jenis dari hidrogen cair adalah 70.8 kg/m^3 dan 70.6 kg/m^3 untuk hidrogen padat. Namun proses kondensasi ini membutuhkan energi yang sangat

besar dimana kebutuhan energi ini membatasi kemungkinan penggunaan sistem penyimpanan hidrogen cair dimana biaya penyimpanan hidrogen tidak menjadi masalah dan gas akan dikonsumsi dalam waktu yang singkat seperti aplikasi pada kendaraan luar angkasa. Sistem penyimpan yang menjanjikan untuk bisa memenuhi persyaratan seperti tersebut di atas adalah dengan menggunakan teknik penyimpan padat (solid-state storage), dimana atom-atom H disimpan di dalam kisi material utamanya. Dalam hal ini hidrogen "disisipkan" pada material tertentu. Metoda solid storage ini dipicu oleh kenyataan bahwa jika menyimpan hidrogen dalam bentuk gas harus dalam bentuk tabung dengan tekanan tinggi (700 bar (4.4 MJ/L)) yang secara safety sangat tidak efisien. Sementara itu, jika disimpan dalam bentuk cair, suhu harus tetap stabil pada -253°C (8 MJ/L). Guna mencari solusi atas permasalahan di atas maka beberapa upaya yang saat ini telah dilakukan untuk mendapat material terutama yang berskala nanometer dan membentuk komposit dan memenuhi persyaratan bagi penyimpan hidrogen padat.

Tabel 2.1 Metode Penyimpanan Hidrogen

Storage method	ρ_m [mass%]	ρ_v [kg H ₂ m ⁻³]	T [°C]	p [bar]	Phenomena and remarks
High pressure gas cylinders	13	<40	RT	800	Compressed gas (molecular H ₂) in light weight composite cylinders (tensile strength of the material is 2000 MPa)
Liquid hydrogen in cryogenic tanks	size dependent	70.8	-252	1	Liquid hydrogen (molecular H ₂), continuous loss of a few % per day of hydrogen at RT
Adsorbed hydrogen	≈ 2	20	-80	100	Physisorption (molecular H ₂) on materials e.g. carbon with a very large specific surface area, fully reversible
Absorbed on interstitial sites in a host metal	≈ 2	150	RT	1	Hydrogen (atomic H) intercalation in host metals, metallic hydrides working at RT are fully reversible

Pada tabel diatas dapat dilihat bahwa penyimpanan hidrogen dalam bentuk gas memiliki presentase massa yang paling tinggi dalam suhu ruangan, namun dengan konsekuensi tekanan yang sangat tinggi (800 bar). Sedangkan hidrogen yang berada pada kisi kisi metal memiliki temperatur rendah, tekanan rendah, serta persen massa yang cukup rendah sehingga banyak penelitian yang dilakukan untuk meningkatkan kapasitas penyimpanan

hidrogen dengan temperatur dan tekanan yang rendah. Terdapat beberapa hal yang perlu diketahui dari teknologi penyimpanan hidrogen seperti bagaimana hidrogen berinteraksi dengan permukaan logam, beserta material yang pada umumnya digunakan sebagai tangki penyimpan hidrogen yang akan dijelaskan dibawah.

2.3. Material Penyimpan Hidrogen

Beberapa jenis logam diyakini memiliki kemampuan menyerap hidrogen dalam jumlah besar. Material-material ini memiliki potensi untuk diaplikasikan pada sistem penyimpan energi hidrogen. Dalam banyak kasus masalah utama adalah sulitnya hidrogen pada saat memasuki material. Tangki bertekanan tinggi yang modern dibuat dari material komposit serat karbon (carbon-fibre-reinforced composite). Tangki modern ini diisi gas bertekanan hingga 600 bar dan biasanya diisi hanya sampai 450 bar untuk keperluan komersial. Begitupun tangki ini butuh pelapis bagian dalam yang sangat khusus (terbuat dari polimer) untuk menghindari bahaya tekanan gas yang sangat tinggi itu. Konsekuensinya, harus digunakan tangki tambahan dari aluminium di dalamnya untuk menjaga interaksi internal antara serat karbon dengan pelapisnya sehingga mengurangi sisi ekonomis dari tangki penyimpanan gas ini.

Perbandingan dari segi fisik berbagai tipe kontainer hidrogen dapat dilihat pada Gambar 2 yang menunjukkan perbandingan antara tangki penyimpan berbasis gas bertekanan tinggi, lalu tangki gas cair, dan tangki hidrogen padat (solid hydrogen storage). Alternatif yang disebut terakhir saat ini sangat aktif diteliti, karena dipandang dari sisi keamanan lebih menguntungkan. Dari gambar jelas ditunjukkan bahwa bila menggunakan penyimpan berbasis gas, dimensi tangki sangat besar sehingga secara tidak langsung kendaraan akan bertambah berat karena dimensi kendaraan akan bertambah. Dalam bentuk cair, sedikit lebih moderat, dimana ukuran tangki dapat direduksi.

Begitupun, menilik temperatur rendah yang dibutuhkan untuk operasionalnya, sistem penyimpanan tipe ini menjadi sulit untuk diaplikasikan secara masal. Akhirnya, sistem penyimpanan padat dimana hidrogen disisipkan ke dalam material/logam tertentu menjadi pilihan paling realistis.



Gambar 2.2 Perbandingan kontainer hidrogen dari segi ukuran fisik

2.3.1. Hydrogen container berbasis logam hidrida

Logam hidrida merupakan kandidat yang sangat menjanjikan untuk aplikasi penyimpanan hidrogen. Pada saat ini, logam hidrida yang paling sering digunakan adalah material anoda pada baterai nikel-metal hydrides (Ni-MH) komersial yang telah menggantikan baterai nikel-cadmium konvensional pada beragam aplikasi. Terdapat beragam variasi dari baterai Ni-MH yang berada di pasaran dimana ukuran selnya berkisar antara 30 mAh to 250 mAh. Riset dan penelitian kebanyakan menargetkan dua aplikasi yaitu baterai dan penyimpanan hidrogen. Kelebihan utama dari penyimpanan hidrogen pada logam hidrida adalah volumetric density yang tinggi (terkadang melebihi hidrogen cair) dan kemungkinan absorpsi dan deabsorpsi hidrogen dengan perubahan yang kecil pada tekanan hidrogen. Logam Hidrida dapat

didefinisikan sebagai senyawa satu fasa antara host metal dan hidrogen. Logam hidrida dapat dibedakan menjadi ionic atau saline hydrides, kovalen hydrides, dan metallic hydrides. Hampir semua hidrida yang dapat digunakan untuk penyimpanan hidrogen bersifat logam. Metallic hydrides dibentuk oleh logam transisi termasuk rare earth dan actinide series. Pada hidrida ini, hidrogen akan bersifat seperti metal dan membentuk ikatan logam yang memiliki konduktivitas listrik dan termal yang tinggi [7].

2.3.2 Logam Hidrida (Metal Hydride)

Reaksi antara intermetallic alloy AB_z dengan gas hidrogen secara umum adalah sebagai berikut



dimana Q adalah panas yang dilepas ketika proses absorpsi hidrogen. Pada umumnya logam A membentuk hidrida binary yang stabil apabila A adalah logam transisi awal, rare-earth metal, atau Mg. Logam B seperti Ni, Co, Cr, Fe, Mn, atau Al tidak membentuk hidrida yang stabil, walaupun logam B dapat membantu proses disosiasi molekul H_2 . Tabel 2 mendeskripsikan representasi hidrida dari lima keluarga logam hidrida (A , A_2B , AB , AB_2 , dan AB_5) yang telah menunjukkan sifat sifat paling menjanjikan untuk teknologi penyimpanan hidrogen. Panas ΔQ yang dilepaskan pada saat absorpsi biasanya dikarakterisasikan oleh parameter entalpi ($\Delta H_{plateau}$) yang ditentukan oleh plot van't Hoff dari tekanan equilibrium ditengah daerah plateau pada pressure–composition–temperature (PCT) isotherms. Untuk dapat dipertimbangkan sebagai penyimpanan bahan bakar pada kendaraan bermotor baik yang ditenagai oleh fuel cell maupun mesin pembakaran dalam, hidrida tersebut harus menyediakan gas hidrogen pada tekanan 1-10 bar dengan range suhu antara 270 K hingga 360 K dan 600 K, tergantung berdasarkan sumber panas yang tersedia. Hidrida $ZrNi$, $ZrMn_2$, atau Logam paduan Mg tidak sesuai untuk aplikasi tersebut. Walaupun logam paduan lain yang ada pada tabel

dibawah dapat menyediakan hidrogen pada tekanan dan temperatur yang diinginkan, namun kapasitas reversibel nya hanya dibawah 2wt% yang akan menyebabkan tambahan massa berlebihan pada kendaraan bermotor.

Tabel 2.2 Properti Utama Logam Hidrida untuk Aplikasi Fase Gas

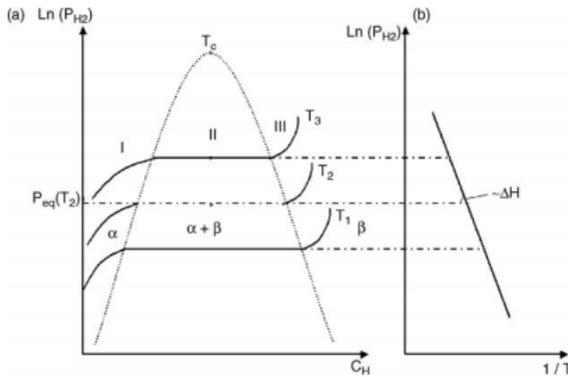
Alloy Type	Crystal Structure Type	Hydride Phase	Maximum H Capacity (wt%)	Reversible H Capacity (wt%)	P_{des}^a at 298 K (bar)	T (K) for 1.013 bar P_{des}	$-\Delta H_{plateau}^b$ (kJ/mol H ₂)	Comments
A	A3 (hP2)	MgH ₂	7.66	<7.0	$\sim 10^{-4}$	552	74.5	Ni improves kinetics
A	A2 (cI2)	VH ₂	3.81	1.9	2.1	285	40.1	Two plateaus, difficult to activate
A ₃ B	C ₄ (hP18)	Mg ₂ NiH ₄	3.59	3.3	$\sim 10^{-4}$	528	64.5	Not metallic, very slow kinetics for T < 500 K
AB	B2 (oP2)	TiFeH ₂	1.89	1.5	4.1	265	28.1	Two plateaus, hard to activate
AB	Bf (oC8)	ZrNiH ₃	1.96	1.1	$\sim 5 \times 10^{-6}$	573	68.6	Two plateaus, fast kinetics
AB ₃	C14 (hP12)	TiMn _{1.4} V _{0.12} H _{3.4}	2.15	1.1	3.6	268	28.6	V makes it more expensive
AB ₃	C14 (hP12)	ZrMn ₂ H _{3.8}	1.77	0.9	0.001	440	53.2	Powder is highly pyrophoric
AB ₃	D _{3d} (hP6)	LaNi ₅ H _{6.5}	1.49	1.28	1.8	285	30.8	Fast kinetics, but degrades for T > 350 K
AB ₃	D _{3d} (hP6)	LaNi _{4.8} Sn _{0.2} H _{6.0}	1.40	1.24	0.5	312	32.8	Slow degradation for T < 550 K

2.4. Hidrogenasi Logam Hidrida

Hidrogenasi adalah proses masuknya senyawa hidrogen yang berbentuk gas ke lokasi interstisi pada logam. Ketika terekspos dengan hidrogen, logam yang dapat membentuk hidrida akan membentuk logam hidrida dengan reaksi sebagai berikut:



dimana Q merupakan panas yang dihasilkan dari reaksi pembentukan hidrida. Aspek termodinamik dari reaksi ini dapat dideskripsikan oleh pressure-composition isotherms (PCI).



Gambar 2.3 Pressure Composition Isotherm

Pada konsentrasi rendah di daerah I ($x \ll 1$) hidrogen akan mula-mula terlarut pada kisi kisi logam dan membentuk fase solid solution (fase α). Hidrogen kemudian akan terdistribusi secara acak pada kisi-kisi logam induk dan konsentrasi akan bervariasi secara perlahan dengan perubahan suhu. Fase α memiliki struktur kristal yang sama dengan logam yang belum dipadukan. Seiring meningkatnya tekanan hidrogen, konsentrasi juga turut meningkat hingga interaksi tarik-menarik H-H menjadi penting. Pada titik ini (daerah II), nukleasi dari fase dengan konsentrasi yang lebih tinggi (fase β) mulai terjadi. Sistem ini sekarang memiliki 3 fase yaitu α , β , dan hidrogen dalam bentuk gas, beserta dua komponen yaitu logam dan hidrogen. Pada daerah ini konsentrasi dari hidrogen akan meningkat dengan tekanan konstan yang ditunjukkan dengan adanya plateau pada grafik. Ketika fase α sudah sepenuhnya menghilang, hidrogen kemudian akan memasuki solid solution di fase β dan tekanan hidrogen akan naik kembali seiring dengan kenaikan konsentrasi hidrogen (daerah III). Tekanan equilibrium P_{eq} pada saaf transformasi fase $\alpha \rightarrow \beta$ didapatkan melalui van't Hoff law:

$$\ln P_{eq} = \frac{\Delta H}{RT} - \frac{\Delta S}{R} \dots \dots \dots (3)$$

dimana ΔH adalah entalpi dan ΔS adalah entropi dari transisi fase $\alpha \rightarrow \beta$. Secara umum perubahan entropi diberikan oleh hilangnya entropi standar ketika gas hidrogen memasuki kisi-kisi logam. ΔS dapat diasumsikan sebagai konstan dimana pada saat proses absorpsi ΔS bernilai 110 KJ/mol dan pada saat desorpsi bernilai 115 KJ/mol [8].

2.5. Logam Hidrida AB_2

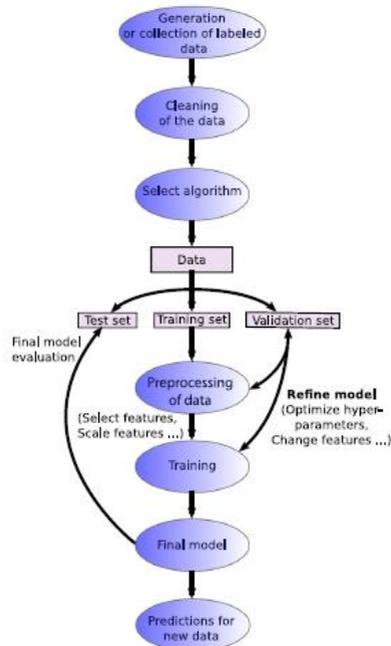
Logam hidrida AB_2 memiliki kapasitas penyimpanan hidrogen yang lebih besar dibandingkan dengan keluarga logam hidrida paduan AB_5 dikarenakan penggunaan logam transisi yang relatif lebih ringan. Senyawa AB_2 dapat dibentuk dari kombinasi berbahaya elemen berbeda. Sandrock [9] berpendapat bahwa elemen A berasal dari grup 4 (Ti, Zr, Hf) atau lanthanoids (La, Ce, Pr, dll.), sedangkan elemen B dapat berasal dari logam transisi maupun non transisi dengan preferensi V, Cr, Mn, dan Fe. Feng et al. [10] menyimpulkan bahwa elemen yang dapat membentuk AB_2 adalah $A = Mg, Zr, Ti$, dan $B = V, Cr, Mn, Ni$. Sebagai contoh kita dapat melihat La dan Ni, yang bergabung untuk memproduksi logam hidrida yang diketahui banyak orang yaitu $LaNi_5$. La akan membentuk hidrida yang sangat stabil, LaH_2 , dengan $\Delta H = -104$ KJ/mol H, sedangkan Ni membentuk hidrida yang sangat tidak stabil dengan $\Delta H = -3$ KJ/mol H. Hidrida yang sangat stabil tidak bagus untuk hidrogen storage karena hal ini ditranslasikan ke temperatur operasi yang sangat tinggi. Oleh karena itu perlu dilakukan pemaduan untuk meraih temperatur operasi serta sifat-sifat lain yang paling tepat untuk aplikasi yang diinginkan. Setiap atom logam pada paduan AB_2 dapat menampung 2 atom hidrogen ($H/M=2$) dimana kapasitas hidrogen maksimum dari AB_2 adalah (~ 3.8 wt%) dan penelitian sebelumnya melaporkan bahwa AB_2 berbasis logam transisi memiliki kapasitas reversible hidrogen sebesar 2.4 wt% [11]. Hal ini tentu saja belum memenuhi batas maksimum yang ditargetkan badan energi dunia sebesar 5 wt%. Namun apabila dibandingkan dengan paduan Mg yang dapat mencapai 7.6 wt %, AB_2 memiliki kelebihan yang cukup banyak

dimana AB_2 memiliki temperatur operasi dibawah $100^\circ C$ dan kinetik yang tinggi, sedangkan paduan Mg beroperasi di temperatur yang sangat tinggi ($300^\circ C$) serta waktu penyerapan hidrogen minimal 60 menit. Oleh karena itu AB_2 memiliki keunggulan dibandingkan logam hidrida paduan yang lain. Riset yang ada saat ini dilakukan untuk memperbaiki sifat-sifat AB_2 tersebut. Misalnya studi pengaruh variasi konsentrasi tiap elemen paduan serta teknik-teknik preparasi terhadap sifat AB_2 . Senyawa AB_2 berbasis logam transisi yang disintesa menggunakan teknik induksi konvensional dan teknik arc-melting biasanya mengandung campuran fase hidrogen C14 atau C15 bersama dengan fase-fase secondary seperti Zr_7Ni_{10} , Zr_9Ni_{11} , $TiNi$, atau $ZrNi$. Fasa C14 dan C15 dapat memberikan kontribusi yang berbeda terhadap sifat-sifat penyimpanan hydrogen pada logam hidrida. Logam hidrida dengan fasa C15 yang lebih dominan diketahui dapat melepaskan hydrogen dengan lebih mudah pada saat proses desorpsi, sedangkan logam hidrida dengan fasa C14 yang lebih dominan dapat menyimpan hydrogen dengan kapasitas yang lebih besar. Dalam pembuatan logam hidrida, diinginkan paduan yang dapat memberikan energy pembentukan serendah mungkin serta kapasitas hydrogen yang setinggi mungkin. Senyawa AB_2 sudah pernah digunakan sebagai material penyimpan hydrogen reversibel. Pada tahun 1908-an armada Daimler vans and automobiles beroperasi menggunakan tangki hydrogen berbasis AB_2 yang mengandung $Ti_{0.98}Zr_{0.02}Cr_{0.05}V_{0.43}Fe_{0.09}Mn_{1.5}$. Senyawa intermetalik ini memiliki kinetik yang sangat cepat serta long term cyclic stability yang bagus. Long term cycling stability dari suatu material didefinisikan sebagai kemampuan material tersebut untuk memelihara kapasitas penyimpanan reversible pada proses charging dan discharging hydrogen yang berulang-ulang [12].

2.6. Penerapan Metode *Machine Learning* pada Ilmu Material

2.6.1. Prinsip dasar *Machine Learning*

Algoritma *machine learning* bertujuan untuk mengoptimasi performansi dari suatu tugas dengan menggunakan contoh atau pengalaman yang telah terjadi di masa lalu [13]. *Machine learning* dapat dibagi menjadi tiga kategori yaitu *supervised learning* atau pembelajaran yang diawasi, *unsupervised learning* atau pembelajaran yang tidak diawasi, serta *reinforcement learning* atau pembelajaran penguatan. *Supervised learning* memiliki prinsip yang sama dengan data *fitting* dimana program akan mencoba untuk mencari suatu fungsi yang tidak diketahui guna menghubungkan input-input yang ada untuk menghasilkan suatu output yang belum diketahui. Hasil tersebut diestimasi menggunakan ekstrapolasi dari pola pola yang ditemukan pada data berlabel yang digunakan untuk melakukan *training* pada program tersebut. *Unsupervised learning* akan lebih berfokus kepada menemukan pola dari data-data yang belum memiliki label seperti mengelompokkan sample. Sedangkan *reinforcement learning* adalah proses analisa masalah untuk menemukan suatu hasil yang paling optimal. Diantara *supervised* dan *unsupervised learning* terdapat *semi-supervised learning* dimana algoritma diberikan data berlabel maupun tidak berlabel. *Supervised learning* merupakan metode yang paling sering digunakan pada ilmu material. Adapun workflow dari *supervised learning* dapat dilihat pada gambar dibawah ini.



Gambar 2.4 Workflow Supervised Learning

Proses dimulai saat seseorang memilih suatu *subset* dari populasi yang relevan dimana nilai-nilai sudah diketahui atau membuat suatu data set baru apabila diperlukan. Proses ini akan didampingi dengan pemilihan algoritma *machine learning* yang akan digunakan untuk fitting semua data. Sebagian besar pekerjaan akan terdiri dari pembuatan, penemuan, dan pembersihan dari data untuk memastikan bahwa data ini konsisten dan akurat. Selain itu perlu properti dari sistem seperti *input* dari model juga perlu dipetakan agar bisa menyesuaikan dengan algoritma yang dipilih. Hal ini bertujuan untuk menerjemahkan informasi mentahan menjadi suatu fitur tertentu yang digunakan sebagai input dari algoritma. Setelah proses tersebut selesai, model akan dilatih dengan mengoptimasi performansi nya, yang biasanya diukur menggunakan suatu fungsi harga. Untuk menghindari *overfitting*

dan *underfitting*, perlu dilakukan monitoring pada saat training untuk menghindari training error maupun error dari validation set. Sebelum model siap untuk digunakan, model tersebut harus dievaluasi pada data sebelumnya yang belum digunakan untuk mengestimasi kemampuan model untuk melakukan generalisasi dan ekstrapolasi. Evaluasi dapat dilakukan menggunakan beberapa metode seperti *over k-fold cross-validation*, *leave-one-out cross-validation*, *Monte carlo cross-validation*, dan *leave-one-cluster-out cross validation*.

2.6.2. Database

Machine learning pada ilmu material pada umumnya menggunakan *supervised learning*. Kesuksesan metode tersebut bergantung kepada jumlah serta kualitas dari data yang tersedia. Hal ini merupakan salah satu dari tantangan utama dari penerapan machine learning pada ilmu material. Hal ini akan menjadi sulit ketika data hanya dapat diperoleh dari eksperimen yang berbiaya mahal seperti temperatur kritis dari material superkonduktor. Oleh karena itu database seperti penelitian terdahulu dan archive nasional yang berisi informasi properti dari material yang sudah diketahui akan sangat berguna untuk menentukan kesuksesan dari metode machine learning ini. Selain itu juga database harus memenuhi empat syarat yaitu: data harus mudah ditemukan, mudah diakses, mudah diinterpretasikan, serta mudah digunakan kembali. Pada umumnya hasil negative dari model tidak akan digunakan serta diterbitkan, namun sebenarnya data negative juga sangat penting sama dengan data positive. Oleh karena itulah, pada akhirnya setiap data baik positif maupun negatif akan sangat penting untuk menentukan kesuksesan penerapan machine learning pada ilmu material.

2.6.3. Fitur

Hal penting yang harus ada pada algoritma *machine learning* adalah representasi data dengan bentuk yang sesuai. Fitur atau variable yang ada pada ilmu material harus dapat menangkap

berbagai informasi relevan yang penting untuk menentukan perbedaan dari berbagai *atomic* atau *crystal environment*. Proses ini dinamakan sebagai ekstraksi fitur yang bisa dilakukan dengan cara yang sederhana seperti menentukan nomor atom, dapat juga berlangsung dengan cara yang kompleks yang melibatkan transformasi, dan dapat juga memerlukan agregasi berdasarkan statistik. Berapa banyak jumlah *processing* yang diperlukan juga sangat bergantung kepada algoritma yang digunakan. Fitur yang paling sering digunakan pada ilmu material adalah fitur yang berhubungan dengan *fitting* dari energi potensial pada permukaan seperti posisi atom yang dapat dideskripsikan dengan koordinat kartesius.

2.6.4. Algoritma Multivariate Regression

Multiple regression merupakan sebuah teknik untuk menganalisa hubungan antara dependent variabel y dan satu atau lebih independent variabel x_1, x_2, \dots, x_k . Tujuan dari multiple linear regression adalah untuk menghasilkan prediksi (prediction) nilai dari suatu variabel dependen berdasarkan nilai-nilai variabel independen. analisis regresi adalah untuk mengidentifikasi sebuah fungsi yang menjelaskan sedekat mungkin hubungan antara variabel dependent sehingga dapat diprediksi dengan menggunakan rentang nilai variabel independent.

Untuk memprediksi nilai dari variabel Y , akan lebih baik apabila ikut memperhitungkan variabel-variabel lain yang ikut mempengaruhi Y . Dengan demikian ada hubungan antara satu variabel tidak bebas (dependent variable) Y dengan beberapa variabel lain yang bebas (independent variable) X_1, X_2, \dots, X_k . Hubungan antara sebuah variabel tak bebas (dependent variable) dengan dua buah atau lebih variabel bebas (independent variable) dalam bentuk regresi disebut dengan Regresi Linear Ganda. Untuk meramalkan Y , apabila semua nilai variabel bebas diketahui, maka dapat digunakan persamaan regresi linear berganda. Hubungan antara Y dan X_1, X_2, \dots, X_k yang sebenarnya adalah

$$Y_i = a + b_1X_1 + b_2X_2 + \dots + b_iX_i \dots\dots\dots(4)$$

dengan

Y_i : Variabel dependent (Nilai yang akan diprediksi)

a : Konstanta

b : Bobot(koefisien) regresi untuk variabel independent ke- i

X_i : Variabel independent ke- i

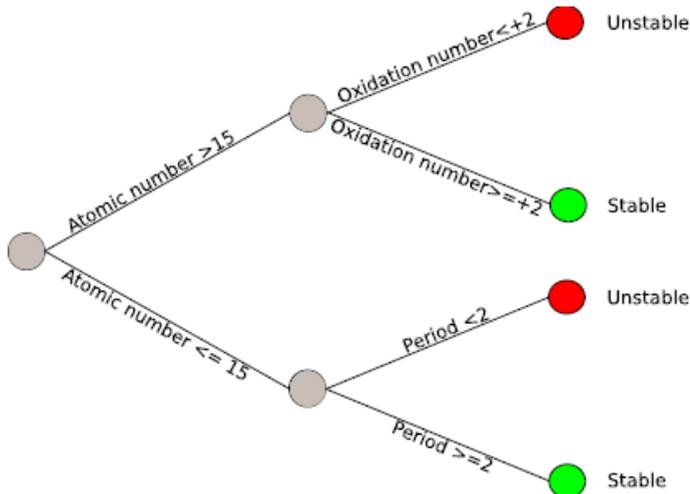
Langkah-langkah analisis regresi linier berganda adalah sebagai berikut:

1. Mengumpulkan sampel data untuk setiap variabel y , x_1 , x_2, \dots, x_k
2. Membuat hipotesis untuk bentuk model y Hal ini melibatkan proses pemilihan variabel predictor yang akan dimasukkan kedalam model.
3. Melakukan uji asumsi tingkat signifikansi σ dan modifikasi model regresi jika diperlukan.
4. Jika model sudah dianggap signifikan, maka model tersebut dapat digunakan untuk memprediksi atau meramalkan nilai variabel respon, melihat seberapa besar pengaruh suatu variabel prediktor terhadap variabel respon, dan membuat kesimpulan lainnya.

2.6.7. Algoritma Decision Tree

Metode *decision tree* atau pohon keputusan adalah salah satu metode regresi dan klasifikasi yang sering digunakan. Metode pohon keputusan mengubah suatu fakta yang besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan tersebut dapat dengan mudah diinterpretasikan oleh manusia. Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah variable input dengan sebuah variable target [14]. Pada decision tree, variabel akan dipisah menjadi dua atau lebih set homogenus. Hal ini dilakukan berdasarkan variabel independen yang paling signifikan untuk membuat grup-grup yang

berbeda. Contoh penggunaan decision tree untuk menentukan apakah suatu material itu stabil adalah sebagai berikut:



Gambar 2.5 Penentuan Kestabilan Atom menggunakan Decision Tree

Pada umumnya, metode decision tree memiliki karakteristik yang cepat dan sederhana untuk di-*training* selain itu decision tree juga cocok untuk dataset yang berukuran besar. Decision tree mampu membreak down proses pengambilan keputusan yang kompleks menjadi lebih simpel sehingga pengambil keputusan akan dapat menginterpretasikan solusi dari permasalahan dengan lebih baik.

Decision tree juga disebut sebagai diagram alir yang berbentuk seperti struktur pohon yang mana setiap internal node menyatakan pengujian terhadap suatu atribut, setiap cabang menyatakan output dari pengujian tersebut dan node daun (leaf node) menyatakan distribusi kelas. Node yang paling atas disebut node akar (root node). Jalir pengujian data adalah pertama, semua data harus melalui root node dan terakhir adalah melalui leaf node

yang akan menyimpulkan prediksi kelas bagi data tersebut. Atribut data harus berupa data kategorik, bila kontinu maka atribut harus didiskretisasi terlebih dahulu.

Data dalam pohon keputusan biasanya dinyatakan dalam bentuk tabel dengan atribut dan record. Atribut menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan *Tree*. Misalkan untuk menentukan main tenis, kriteria yang diperhatikan adalah cuaca, angin dan temperatur. Salah satu atribut merupakan atribut yang menyatakan data solusi per-item data yang disebut dengan target atribut. Atribut memiliki nilai-nilai yang dinamakan dengan instance.

Proses pada pohon keputusan adalah mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi rule, dan menyederhanakan rule. Langkah pertama yang dilakukan dalam pembuatan pohon keputusan yaitu menghitung nilai *Entropy* total dari jumlah data yang dijadikan sampel, selanjutnya yaitu mengelompokkan variabel dan menghitung nilai *Gain* pada tiap atribut. Setelah dihitung menggunakan rumus algoritma C4.5 maka atribut yang mempunyai nilai *Gain* tertinggi akan menjadi akar dan atribut lainnya menjadi cabang, kemudian dari cabang akan dihitung kembali atribut apa lagi yang mempunyai nilai *Gain* tertinggi. Tahapan perhitungan akan berulang secara continue sampai beberapa tingkatan hingga mencapai nilai akhir yaitu keputusan “Yes” dan “No”.

Untuk memilih atribut sebagai akar, didasarkan pada nilai *Gain* tertinggi dari atribut-atribut yang ada. Untuk menghitung *Gain* digunakan rumus seperti tertera dalam persamaan berikut:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \dots \dots \dots (5)$$

Keterangan:

S: himpunan kasus

A: atribut
 n: jumlah partisi atribut A
 $|S_i|$: jumlah kasus pada partisi ke-i
 $|S|$: jumlah kasus dalam S

Setelah mendapatkan nilai *Gain*, ada satu hal lagi yang perlu dilakukan perhitungan yaitu mencari nilai *Entropy*. *Entropy* digunakan untuk menentukan seberapa informatif sebuah input atribut untuk menghasilkan output atribut. Rumus dasar dari *Entropy* tersebut adalah sebagai berikut:

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i \dots \dots \dots (6)$$

Keterangan:

S: himpunan kasus
 n: jumlah partisi S 14
 pi: proporsi dari Si terhadap S

2.6.8. Algoritma Random Forest

Random Forest merupakan sebuah metode ensemble. Metode ensemble merupakan cara untuk meningkatkan akurasi metode klasifikasi dengan cara mengkombinasikan metode klasifikasi [15]. Random Forest diawali dengan teknik dasar data mining yaitu decision tree. Pada decision tree input dimasukan pada bagian atas (root) kemudian turun kebagian bawah (leaf) untuk menentukan data tersebut termasuk kelas apa. Random forest adalah pengklasifikasi yang terdiri dari kumpulan pengklasifikasi pohon terstruktur dimana masing-masing pohon melemparkan unit suara untuk kelas paling populer di input x. Dengan kata lain Random Forest terdiri dari sekumpulan decision tree (pohon keputusan), dimana kumpulan decision tree tersebut digunakan untuk mengklasifikasi data ke suatu kelas.

Random forest merupakan metode klasifikasi yang supervised. Sesuai dengan namanya, metode ini menciptakan

sebuah hutan (forest) dengan sejumlah pohon (tree). Secara umum, semakin banyak pohon (tree) pada sebuah hutan (forest) maka semakin kuat juga hutan tersebut terlihat. Pada kasus yang sama, semakin banyak tree, maka semakin besar pula akurasi yang didapatkan.

Decision Tree akan menggunakan information gain dan gini index untuk perhitungan dalam menentukan root node dan rule. Sama halnya dengan Random Forest yang akan menggunakan information gain dan gini index untuk perhitungan dalam membangun tree, hanya saja Random Forest akan membangun lebih dari satu tree. Masing-masing tree dibangun menggunakan set data dengan atribut yang diambil secara acak dari data training. Dengan kata lain setiap tree akan bergantung pada nilai dari sampel vektor yang independen dengan distribusi yang sama pada setiap tree. Selama proses klasifikasi setiap tree akan memberikan voting kelas yang paling populer.

2.6.9. Interpretabilitas dari Metode *Machine Learning*

Terdapat berbagai kritik dari para ahli terkait teknik machine learning seperti algoritma kotak hitam dimana kita tidak mengetahui secara pasti proses apa yang dilakukan oleh suatu model ketika kita memberikan suatu input. Kekhawatiran utama ialah bahwa suatu model yang tidak dibuat berdasarkan hukum fisika, melainkan hanya berasal dari suatu set data, akan gagal pada kasus-kasus tertentu yang tidak diharapkan oleh manusia walaupun mungkin model tersebut akan sukses pada rata-rata percobaan yang dilakukan. Hal ini akan menjadi berbahaya pada aplikasi dimana satu kali kegagalan berharga sangat mahal atau bisa berakibat fatal (seperti pada aplikasi di dunia obat-obatan). Kasus-kasus tertentu ini hanya akan dapat diproses secara benar apabila pengguna mengetahui secara pasti hubungan sebab akibat dari input dan output. Terdapat beberapa konsep interpretabilitas yang perlu diketahui yaitu transparansi dan penjelasan post hoc. Transparansi dapat dibagi menjadi tiga konsep. Konsep pertama adalah

simultability yaitu sebagian gagasan harus bersifat subjektif dan manusia harus dapat menelusuri kembali jejak perhitungan dari model. Konsep kedua adalah *decomposability* yang berarti apakah bagian dari model seperti input, parameter, dan perhitungan dapat diinterpretasikan secara intuitif. Konsep yang ketiga adalah transparansi algoritma yang mempertimbangkan pemahaman pengguna terhadap error yang mungkin terjadi. Sedangkan penjelasan post hoc mempertimbangkan kemungkinan untuk mendapatkan suatu informasi tambahan dari model tersebut. Setiap konsep interpretabilitas ini memiliki prioritas yang berbeda tergantung pada dataset dan tujuan dari riset. *Simultability* biasanya tidak ada pada ilmu material, yang berarti bahwa *simultability* sekarang harus menjadi pertimbangan dalam penggunaan machine learning pada ilmu material. Salah satu cara untuk meningkatkan kepercayaan pengguna terhadap model adalah dengan meningkatkan penggunaan metode analisa post hoc.

2.7. Penelitian Terdahulu

Metode machine learning telah terbukti sukses dalam memprediksi berbagai properti dari material. Suwarno dan Shahab [2] pada tahun 2004 melakukan penelitian prediksi derajat presipitasi karbida krom pada baja tahan karat austenitik menggunakan metode neural network atau jaringan syaraf tiruan (JST). Presipitasi karbida pada batas butir menyebabkan baja menjadi lebih rentan terhadap korosi batas butir. Derajat presipitasi dari karbida ini dapat dideteksi dengan melakukan pengamatan makrografi pada spesimen yang dipanasi secara lokal, dimana hasil pengamatan makrografi yang konsisten dijadikan sebagai data untuk menyusun jaringan syaraf tiruan sebagai model yang dapat diunakan untuk memprediksi derajat presipitasi karbida. JST berhasil memprediksi dengan baik derajat presipitasi yang dibuktikan dengan tidak adanya perbedaan yang signifikan pada uji statistik. Selain prediksi derajat presipitasi karbida, metode machine learning juga telah digunakan pada aplikasi logam hidrida. Rahnema et. al. [3] pada 2019 menggunakan metode

machine learning untuk memprediksi properti dari logam hidrida sebagai penyimpanan hidrogen. Properti yang diprediksi adalah kelas material beserta *hydrogen weight percent*. Penelitian yang dilakukan memanfaatkan database yang diberikan oleh Departemen Energi Amerika Serikat dimana properti logam hidrida seperti *hydrogen-weight percent*, *heat of formation*, serta temperatur dan tekanan operasi dianalisa melalui *supervised machine learning* dengan empat metode yaitu *linear regression*, *neural network*, *Bayesian linear regression*, dan *boosted decision tree*. Hasil dari masing-masing algoritma dibandingkan satu dengan yang lain dan ditemukan bahwa *Bossted Decision Tree* memiliki performansi yang paling baik diantara keempat metode tersebut untuk memprediksi properti dari logam hidrida. Hattrick-Simpers, et.al [16] pada tahun 2018 melakukan penelitian untuk memprediksi paduan logam hidrida yang paling potensial untuk digunakan sebagai *hydrogen storage* dengan menggabungkan analisa ekonomi dengan analisa machine learning untuk menemukan logam hidrida dengan harga termurah dan property yang paling menguntungkan. Sama halnya dengan percobaan yang dilakukan Rahnama et.al, Hattrick-Simpers, et.al juga menggunakan database badan energu amerika serikat untuk mendapatkan data komposisi logam hidrida serta sifat-sifat penyimpanan hidrogennya. Witman et.al [17] pada tahun 2020 melakukan analisa terhadap database logam hidrida yang dimiliki oleh HydPARK untuk menganalisa hubungan antara feature importance dengan struktur yang ada pada logam hidrida yang secara umum dapat menentukan termodinamika pembentukan hidrida. Penelitian yang dilakukan witman et.al memberikan hasil berupa model machine learning yang dapat memprediksi paduan baru yang belum pernah diinvestigasi sebelumnya. Berbeda dengan penelitian yang terdahulu, penelitian yang dilakukan oleh penulis akan lebih berfokus kepada logam hidrida AB_2 yang diketahui dapat memberikan sifat-sifat penyimpanan yang lebih menguntungkan dibandingkan dengan logam hidrida tipe lain seperti AB_5 . Penelitian terdahulu lebih berfokus terhadap

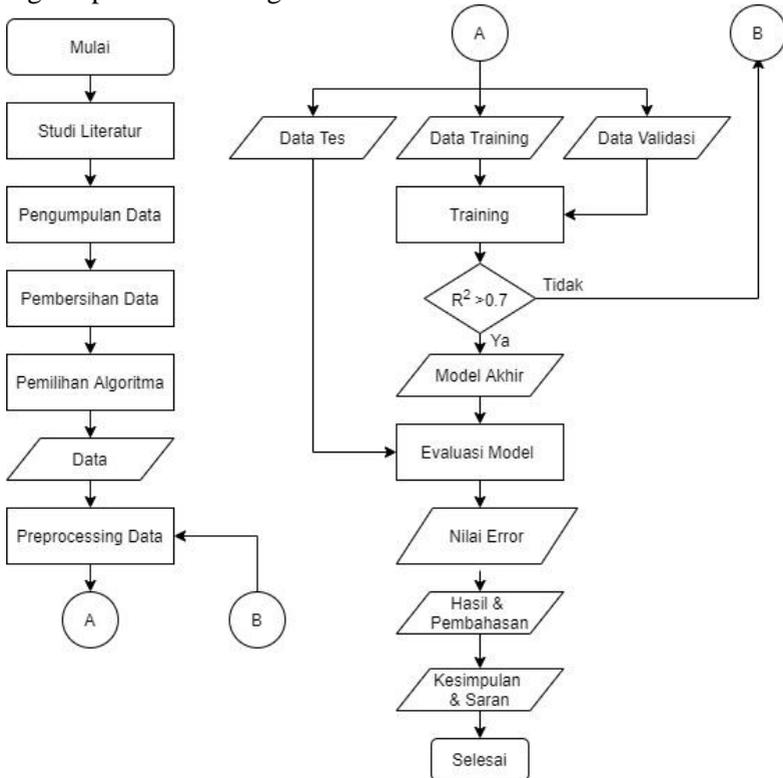
pengklasifikasian paduan yang dimiliki terhadap tipe tipe logam paduan yang ada seperti AB_2 , AB_5 , A_2B , dsb. Pada penelitian ini peneliti ingin berfokus terhadap satu jenis logam hidrida untuk meningkatkan kemampuan prediksi model machine learning. Selain itu, database yang digunakan pada penelitian terdahulu mengandung data data yang berasal dari jurnal yang diterbitkan sejak tahun 1960-an, sedangkan pada penelitian ini akan lebih berfokus terhadap jurnal yang diterbitkan mulai tahun 1998 untuk meningkatkan *reliability* dari dataset yang digunakan.

Halaman ini sengaja dikosongkan

BAB III METODE PENELITIAN

3.1. Diagram Alir Penelitian

Pada penelitian ini akan dibuat model prediksi menggunakan metode machine learning. Metode machine learning yang digunakan adalah supervised learning dengan langkah-langkah penelitian sebagai berikut.

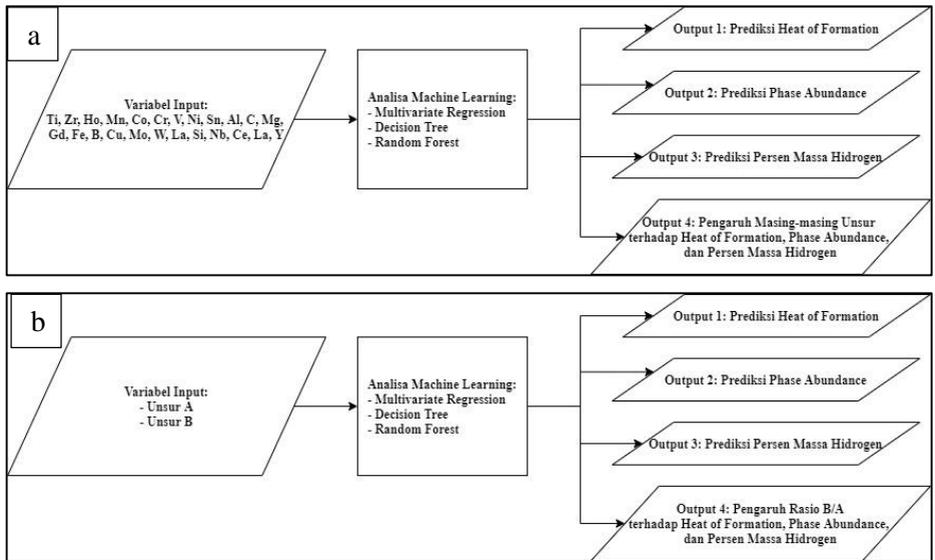


Gambar 3.1 Diagram Alir Penelitian

3.2. Data dan Sumber Data

Prediksi menggunakan metode machine learning membutuhkan data yang cukup banyak. Data tersebut diperoleh dari peneliti terdahulu yang menggunakan variable-variabel tertentu yang sama. Data peneliti terdahulu didapatkan dari 50 jurnal ilmiah yang didapatkan dari *Google Scholars*.

3.3. Desain Eksperimen



Gambar 3.2 Ilustrasi Penelitian Secara Garis Besar untuk (a) Variabel dengan Unsur-Unsur yang Tidak Dikelompokkan, (b) Variabel dengan Unsur-Unsur yang Dikelompokkan Menjadi Unsur A dan B

Terdapat dua variable input yang akan digunakan, beberapa variabel diseleksi dimana komposisi kimia bahan yang menyangkut kandungan Ti, Zr, Ho, Mn, Co, Cr, V, Ni, Sn, C, Mg, Gd, B, Cu, Mo, W, La, Si, Nb, dan Ce digunakan sebagai variabel

input yang pertama seperti pada gambar 3.2. Sedangkan, variable input yang kedua adalah unsur-unsur kimia yang dikelompokkan menjadi elemen A dan B sesuai dengan logam paduan AB_2 dimana A terdiri dari unsur Ti dan Zr, sedangkan B terdiri dari unsur sisa selain Ti dan Zr. Komposisi kimia yang digunakan disini akan dimasukkan dengan satuan mol. Untuk variabel target atau variabel output digunakan data delta entalpi (ΔH) dan delta entropi (ΔS) pada saat material mengalami absorpsi hidrogen, phase abundance C14, serta persen massa hidrogen. Data entalpi dan entropi didapatkan dari tabel yang disediakan pada jurnal ilmiah dengan satuan $KJ/(molH_2.K)$, phase abundance dalam satuan persen (%), dan persen massa hidrogen dalam satuan persen (%). Bagi data selisih entalpi dan selisih entropi yang tidak tersedia dilakukan perhitungan dari nilai yang diambil pada grafik PCT isotherm dengan menggunakan persamaan $P_{eq} = \Delta H/RT - \Delta S/R$ dimana ΔS saat absorpsi dapat diasumsikan $110 KJ/(molH_2K)$ dan $115 KJ/(molH_2K)$ pada saat desorpsi, hal ini dapat dilakukan dengan alasan bahwa ΔS secara teori hanya dipengaruhi oleh perubahan fase hidrogen dari wujud gas menjadi logam hidrida. Dari data peneliti terdahulu didapatkan 313 pasang data yang dapat digunakan untuk mencari model machine learning. Data akan dibagi menjadi tiga kelompok yaitu 70 % pasang data training, 15% pasang data testing, dan 15% pasang data validasi yang dipilih secara random. Perbandingan persen ketiga pasang data tersebut didapatkan dari teori statistika dan machine learning secara umum, selain itu angka 70% pada data training dinilai dapat memberikan informasi yang cukup untuk membangun suatu model, apabila nilai 70% dikurangi maka dikhawatirkan akan menyebabkan model prediksi menjadi tidak akurat, sedangkan apabila nilai 70% dinaikkan maka akan berpengaruh terhadap kurangnya data yang digunakan untuk melakukan validasi dan testing. Data training berfungsi untuk menghitung bobot keterkaitan antara input dan output machine learning, data validasi berfungsi untuk menghentikan proses iterasi ketika terjadi error yang terlalu besar, sedangkan data testing digunakan untuk pengujian model yang

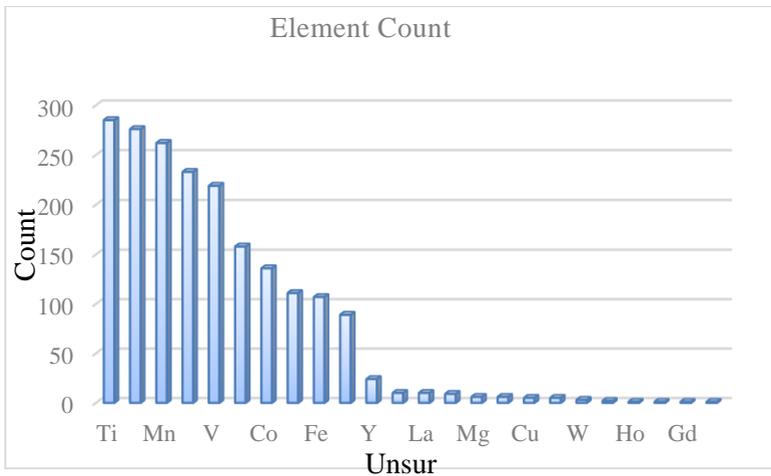
telah mengalami training untuk menemukan nilai error. Data preprocessing dilakukan sebelum memulai training untuk memperkecil kemungkinan terjadinya error, dimana terdapat batas unData ini akan diolah untuk membentuk suatu model dengan menggunakan Microsoft Excel dan Python.

3.4. Pemilihan Algoritma

Terdapat tiga algoritma regresi yang akan digunakan pada prediksi nilai sifat sifat penyimpanan hydrogen yaitu Multivariate Linear Regression, Decision Tree, dan Random Forests. Multivariate Linear Regression dipilih karena merupakan algoritma paling dasar dalam proses analisa machine learning. Decision tree dipilih untuk mengatasi kemungkinan bahwa dataset tersebut bersifat non-linear sehingga dibutuhkan metode selain multivariate linear regression, sedangkan random forest dipilih dengan alasan bahwa model mungkin menghasilkan performansi yang lebih baik ketika terdapat lebih dari satu tree. Algoritma ini telah digunakan pada penelitian terdahulu yang menemukan bahwa ketiga algoritma tersebut memberikan performansi yang baik untuk melakukan regresi pada ilmu material. Selain itu ukuran dataset yang dimiliki tergolong cukup besar sehingga cocok untuk menggunakan ketiga algoritma tersebut. Pengaruh elemen secara individu terhadap sifat penyimpanan hydrogen diteliti menggunakan linear regression dan decision tree, sedangkan pengaruh nilai e/a terhadap sifat penyimpanan hydrogen diteliti menggunakan linear regression.

3.5. Analisa Awal dan Visualisasi Data

Analisa awal dilakukan untuk mendapatkan suatu visualisasi awal dari dataset yang sudah didapatkan untuk memprediksi hasil yang akan didapatkan dari analisa menggunakan machine learning. Dilakukan data count untuk mengetahui prevalensi elemen dari dataset dengan hasil sebagai berikut:



Gambar 3.3 Grafik Visualisasi Prevalensi Elemen pada Dataset

Berdasarkan prevalensi elemen yang ada, indepen awal ini akan dilakukan menggunakan multivariate regression dimana variable independent dengan nilai prevalensi diatas 130 yaitu Ti, Zr, Mn, Cr, V, Ni, dan Co akan masing masing diregresi dengan variable dependen berupa entalpi dan entropi saat 33ndepen, entalpi dan entropi pada saat desorbsi tidak dimasukkan ke dalam 33ndepen awal karena pada berdasarkan hukum konservasi energy kedua tahap ini dapat diasumsikan memiliki entalphi dan entropi yang sama. Angka count 130 didapatkan dari asumsi serta analisa peneliti berdasarkan visualisasi prevalensi elemen yang didapatkan, guna meningkatkan kemampuan prediksi dari model regresi, karena secara teori ketika dataset memiliki nilai 0 yang lebih sedikit maka kemampuan prediksi akan meningkat. Hasil multivariate regression yang dilakukan menggunakan software Microsoft Excel didapatkan sebagai berikut. Statistika Regresi dengan nilai level signifikansi 95 % ($\alpha = 0.05$) dengan hipotesa,

$$H_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0$$

$$H_1 = \text{at least one } \beta_i \neq 0$$

Menmemberikan hasil sebagai berikut,

- a. Regresi komposisi kimia terhadap selisih entalphi absorpsi
 -R square = 0.2555967 (menunjukkan bahwa kemampuan prediksi regresi multivariate adalah 25,5% yang berarti bahwa 25,5% dari entalphi absorpsi dapat dijelaskan menggunakan variable komposisi kimia)
 -Nilai P keseluruhan = 2.67E-16 (Ho ditolak, setidaknya ada satu variable independent yang signifikan)

Tabel 3.1 Koefisien Regresi Linear Prediksi Delta H

No	Unsur	Koefisien	Nilai P	Signifikansi
1	-	114,13	-	-
2	Ti	-83,9	0,393	Tidak
3	Zr	-77,46	-,431	Tidak
4	Mn	-2,667	0,134	Tidak
5	Co	-13,68	0,009	Signifikan
6	Cr	-4,40	0,019	Signifikan
7	V	8,243	0,000128	Signifikan
8	Ni	1,421	0,322	Tidak

Persamaan Regresi:

$$\hat{Y} = 114.13 - 83.9Ti - 77.46Zr - 176.072Ho - 2.667Mn - 13.68Co - 4.40Cr + 8.243V + 1.421 Ni \dots \dots \dots (7)$$

- b. Regresi komposisi kimia terhadap selisih entropi absorpsi
 -R square = 0.122785 (menunjukkan bahwa kemampuan prediksi regresi multivariate adalah 12,2% yang berarti bahwa 12,2% dari entropi absorpsi dapat dijelaskan menggunakan variable komposisi kimia)
 -Nilai P keseluruhan = 9.92E-07 (Ho ditolak, setidaknya ada satu variable independent yang signifikan)

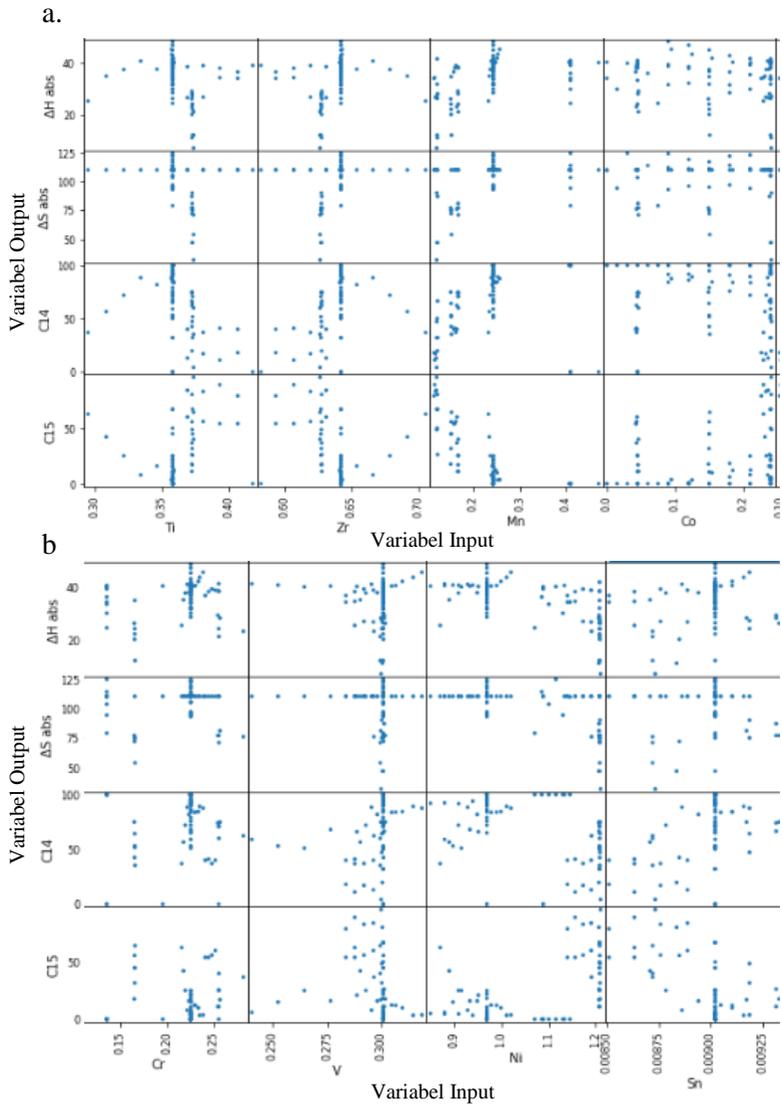
Tabel 3.2 Koefisien Regresi Linear Prediksi Delta S

No	Unsur	Koefisien	Nilai P	Signifikansi
1	-	239,5	-	-
2	Ti	-130,061	0,366	Tidak
3	Zr	-130,097	0,366	Tidak
4	Mn	-4,589	0,077	Tidak
5	Co	-41,134	0,0001	Signifikan
6	Cr	0,503	0,0854	Signifikan
7	V	4,924	0,1137	Tidak
8	Ni	1,3513	0,5199	Tidak

Persamaan Regresi:

$$\hat{Y} = 239.35 - 130.061Ti - 130.097Zr - 4.589Mn - 41.134Co - 0.503Cr + 4.924V + 1.3513 Ni \dots \dots \dots (8)$$

Regresi yang dilakukan menggunakan python jupyter notebook menghasilkan nilai yang sama dengan nilai yang diberikan oleh microsoft excel. Hal ini terjadi karena memang pada dasarnya prinsip regresi linear pada microsoft excel sama dengan python. Namun, nilai 25,5 % dan 12,2% untuk kemampuan prediksi ini tergolong sangat lemah. Oleh karena itu perlu ditemukan penyebab lemahnya kemampuan prediksi ini. Untuk melakukan hal tersebut, dilakukan visualisasi plot dataset variabel input dengan variabel output dan didapatkan hasil sebagai berikut:



Gambar 3.4 Plot Variabel Input dengan Variabel Output untuk a) Ti – Co dan b) Cr - Sn

Berdasarkan plot pada gambar 10 dapat dilihat bahwa terdapat banyak noise yang terjadi pada plot tersebut yang membuat trend pengaruh masing masing variabel input terhadap variabel output menjadi susah untuk dilihat. Sehingga wajar apabila hasil regresi awal menunjukkan kemampuan yang sangat lemah. Noise ini dapat disebabkan karena banyak hal, antara lain: i) 50 Jurnal yang didapatkan berasal dari peneliti serta laboratorium yang berbeda, sehingga terdapat kemungkinan set up percobaan yang berbeda pula. Seperti contohnya terdapat spesimen yang mengalami perlakuan panas dan ball milling. ii) Terdapat banyaknya nilai nol pada dataset tersebut.

Oleh karena itu dilakukan analisa awal tahap kedua yaitu dengan melakukan feature engineering terhadap dataset, dimana dataset yang memiliki nilai 0 tidak akan diikutsertakan dalam regresi linear yang baru. Selain itu dilakukan juga data scaling untuk menyetarakan kemampuan masing masing fitur yang ada, sehingga semua fitur memiliki kesempatan yang sama untuk berkontribusi di proses regresi. Berdasarkan regresi linear tahap kedua, didapatkan hasil sebagai berikut:

a. Regresi komposisi kimia terhadap ΔH absorpsi:

R square = 0.410 (kemampuan prediksi 41%)

Nilai P keseluruhan = 2.46E-10 (H0 ditolak)

$$\hat{Y} = 26.39 + 16.70Ti + 9.68Zr + 16.75Mn + 6.86Co + 50.52Cr - 4.58V - 17.46 Ni \dots \dots \dots (9)$$

b. Regresi komposisi kimia terhadap ΔS absorpsi:

R square = 0.318 (kemampuan prediksi 31.8%)

Nilai P keseluruhan = 3.03E-07 (H0 ditolak)

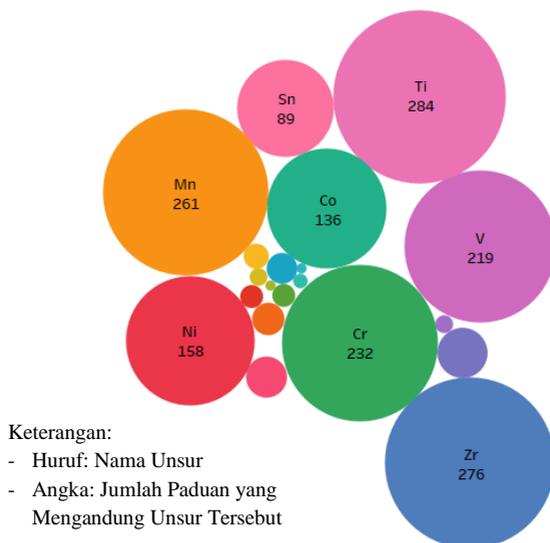
$$\hat{Y} = 77.63 + 35.27Ti + 42.35Zr + 30.401Mn + 18.897Co + 58.51Cr + 30.24V - 43.67 Ni \dots \dots \dots (10)$$

Nilai kemampuan regresi dari terhadap kedua variabel output menunjukkan kenaikan yang cukup signifikan yaitu pada regresi terhadap ΔH yang semula bernilai 25,5% naik menjadi 41% dan pada ΔS yang semula bernilai 12,2% naik menjadi 31,8%. Hal ini menunjukkan bahwasanya sangat penting untuk dilakukan feature engineering ketika mengolah data sebelum dilakukan fitting terhadap model. Pada penelitian ini akan dilakukan data fitting yang lain berupa pengkategorian dataset berdasarkan set up percobaan yang dilakukan.

BAB IV

HASIL DAN PEMBAHASAN

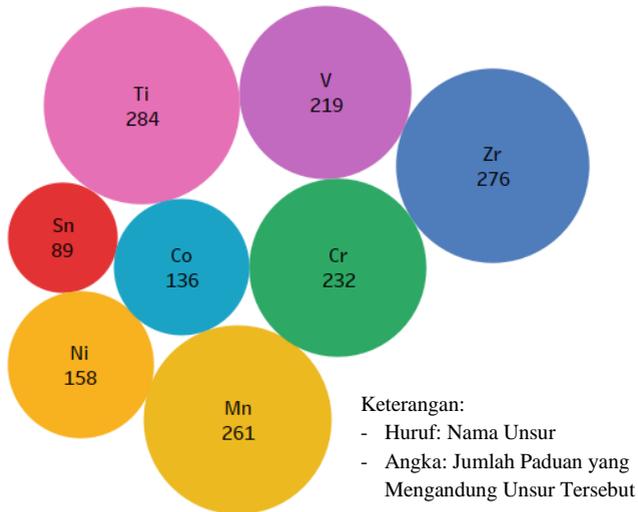
4.1. Data Awal



Gambar 4.1 Prevalensi Elemen pada Dataset Awal

Didapatkan 324 pasang data logam paduan dari 50 jurnal, dimana setiap logam paduan memiliki kandungan unsur yang berbeda-beda. Gambar 4.1 menunjukkan grafik bubble prevalensi elemen yang berada pada seluruh set paduan. Angka yang berada dibawah keterangan jenis elemen menunjukkan jumlah set paduan yang mengandung elemen tersebut. Dapat dilihat bahwa terdapat bubble yang berukuran sangat kecil yang tidak memiliki keterangan jenis elemen unsur. Hal ini menunjukkan bahwa hanya terdapat sedikit logam paduan yang mengandung elemen-elemen tersebut. Keberadaan elemen unsur dengan prevalensi yang sangat rendah ini dapat memberikan pengaruh yang buruk terhadap model

machine learning karena keterbatasan informasi yang diberikan oleh elemen unsur tersebut dapat menyebabkan prediksi yang tidak akurat. Oleh karena itu dilakukan pemilihan variabel input yang akan digunakan pada model dimana diambil 8 unsur yang memiliki prevalensi terbesar yaitu Ti, Zr, Mn, Cr, V, Ni, Co, dan Sn.



Gambar 4.2 Prevalensi Elemen pada Dataset untuk 8 Unsur dengan Prevalensi Terbesar

Gambar 4.2 menunjukkan prevalensi elemen setelah dilakukan penghapusan untuk unsur-unsur yang memiliki prevalensi yang rendah. Gambar menunjukkan bahwa sudah tidak terdapat perbedaan ukuran bubble yang cukup besar. Namun, dari gambar tersebut dapat dilihat bahwa akan ada banyak baris dataset yang memiliki nilai 0 karena tidak semua paduan mengandung unsur-unsur yang ada pada gambar 4.2. Hal ini juga akan menyebabkan menurunnya kemampuan prediksi dari model. Oleh karena itulah dilakukan pengolahan terhadap data awal yang akan dijelaskan pada bagian subbab 4.2.

4.2. Hasil Pengolahan Awal Data

Pengolahan data awal dilakukan untuk mengatasi temuan-temuan pada analisis awal menggunakan multivariate regression seperti banyaknya noise yang dihasilkan oleh set up percobaan yang berbeda-beda, banyaknya *null value*, dan banyaknya data outliers yang dapat menurunkan performa prediksi model *machine learning*. Pengolahan awal data dimulai dengan *feature engineering* dan dilanjutkan dengan *data preprocessing*.

4.2.1. Feature Engineering

Dilakukan *feature engineering* untuk memilih variasi fitur yang memiliki kemampuan prediksi yang paling tinggi, sebagai fitur yang akan digunakan pada pengujian secara keseluruhan. Kedua jenis variabel input yang ada pada penelitian ini akan mendapatkan perlakuan *feature engineering* yang berbeda.

a. Variabel input dengan unsur-unsur elemen yang tidak dikelompokkan

Ti	Zr	Mn	Co	Cr	V	Ni	Sn	$\Delta H (-)$	$\Delta S (-)$	C14	wt%
0.3	0.7	0	0	0	0.5	0	0	26.5	110	100	1.65
0.515	0.485	1.2	0	0.8	0	0	0	33	103	0	1.69
0.515	0.485	1.142857	0	0.761905	0	0	0	30	99	0	1.58
0.515	0.485	1.142857	0.095238	0.761905	0	0	0	29	97	0	1.57
0.515	0.485	1.142857	0	0.761905	0	0.095238	0	31	105	0	1.63
0.358209	0.641791	0.243609	0.240602	0.225564	0.300752	0.968421	0.009023	33.7	105	100	1.74
0.358209	0.641791	0.243609	0.210526	0.225564	0.300752	0.968421	0.009023	35.4	111	100	1.75
0.358209	0.641791	0.243609	0.180451	0.225564	0.300752	0.968421	0.009023	36.4	114	100	1.58
0.358209	0.641791	0.243609	0.150376	0.225564	0.300752	0.968421	0.009023	36.7	111	100	1.49
0.358209	0.641791	0.243609	0.120301	0.225564	0.300752	0.968421	0.009023	37.8	111	100	1.58
0.358209	0.641791	0.243609	0.090226	0.225564	0.300752	0.968421	0.009023	38.7	112	100	1.66
1	0	0.3	0	1.1	0	0	0	16.47	103.75	100	1.53
1	0	0.295567	0	1.083744	0	0	0	16.63	104.17	100	1.715
1	0	0.295567	0	1.083744	0	0	0	19.27	112.11	100	1.629
1	0	0.3	0	1.083744	0	0	0	19.39	111.29	100	1.663
0.142857	0.857143	0.584615	0.153846	0.276923	0	0.984615	0	33.9	110	52	1.68
0.142857	0.857143	0.584615	0.153846	0.276923	0	0.969231	0	35.7	110	44.2	1.63
0.142857	0.857143	0.584615	0.153846	0.276923	0	0.953846	0	34.7	107	41.3	1.62
0.142857	0.857143	0.584615	0.153846	0.276923	0	0.923077	0	32.9	98	44.3	1.6
0.142857	0.857143	0.584615	0.153846	0.276923	0	0.969231	0	36	111	36.5	1.74

Gambar 4.3 Contoh Dataset untuk Variabel Input dengan Unsur-
Unsur yang Tidak Dikelompokkan

Dataset dengan unsur-unsur yang tidak dikelompokkan berguna untuk mengetahui pengaruh masing masing unsur terhadap variabel output pada model serta mendapatkan prediksi model secara keseluruhan. Dimana untuk dataset ini, terdapat enam jenis variasi fitur yang akan diuji yaitu: i) seluruh fitur tanpa dilakukan persuntingan, ii) fitur yang dibagi berdasarkan set up pengujian tanpa menggunakan heat treatment, iii) fitur untuk paduan yang mengalami heat treatment pada set up pengujian, dan iv) fitur untuk senyawa tanpa menyertakan paduan yang memiliki komposisi yang bernilai 0 dimana dilakukan penghapusan untuk senyawa yang memiliki nilai nol seperti pada sel yang diblok menggunakan warna kuning pada gambar 4.3 sehingga dataset yang diuji hanyalah set paduan yang mengandung kedelapan unsur pada variabel input seperti sel yang diblok menggunakan warna hijau. Pengujian variasi fitur ini dilakukan dengan menggunakan multivariate linear regression terhadap seluruh variable target yaitu ΔH absorpsi, ΔS absorbs, phase abundance, dan persen massa hydrogen di software orange python, dengan data train sebanyak 80% dan data tes sebanyak 20 % secara random dan bertingkat. Dari hasil pengujian variasi tersebut didapatkan hasil sebagai berikut.

Tabel 4.1 Nilai Kemampuan Prediksi Seluruh Variable Target
untuk Masing-masing Variasi Fitur.

Variasi Fitur	absorpsi	absorpsi	Phase Abundance	Persen Massa Hidrogen	Skor rata rata
Seluruh fitur tanpa disunting	0.13	0.042	0.317	0.23	0.151
Fitur set up uji dengan	0.684	0.94	0.476	0.22	0.368

heat treatment					
Fitur set up uji tanpa heat treatment	0.257	0.105	0.332	0.445	0.284
Fitur tanpa nilai 0	0.42	0.19	0.86	0.66	0.532

Berdasarkan pengujian yang dilakukan skor rata-rata tertinggi diraih oleh fitur tanpa nilai 0, dan secara berurutan diikuti fitur set up uji dengan heat treatment, fitur set up uji tanpa heat treatment, dan seluruh fitur yang tidak tidak disunting. Pengujian fitur senyawa A dan B secara terpisah mendapatkan skor terendah karena memang pada dasarnya seluruh unsur memiliki pengaruh terhadap variable output sehingga wajar bahwa pada saat senyawa A dan B dipisahkan, skor terendah didapatkan. Pengujian fitur tanpa nilai 0 mendapatkan skor tertinggi diantara semua variasi. Hal ini sesuai dengan teori analisis data bahwa data dengan nilai null yang lebih sedikit akan menghasilkan prediksi yang lebih baik. Oleh karena itu dataset fitur dengan unsur tanpa nilai 0 akan dipilih sebagai dataset yang digunakan untuk variabel dengan unsur-unsur yang tidak dikelompokkan pada pengujian selanjutnya.

b. Variabel input dengan unsur-unsur elemen yang dikelompokkan sebagai unsur A dan B

A	B	$\Delta H (-)$	$\Delta S (-)$	C14	wt%
1.1	2	14.7	92.2	0	0.57
1	2	16.3	89.8	0	1.76
1	2	17.3	91.8	0	1.54
1	2	16.9	89.8	0	0.89
35.7	64	29	77	74	1.3
34.6	65	28	81	74.8	1.44
33.5	66.1	24	77	70.7	1.4
32.2	68.4	23	76	62.3	1.31
31.3	68.4	21	71	60	1.11

Gambar 4.4 Contoh Dataset dengan Unsur-Unsur yang Dikelompokkan Menjadi Unsur A dan B

Variabel dengan unsur-unsur elemen yang dikelompokkan menjadi unsur A dan B berguna untuk mengetahui pengaruh paduan secara stoichiometry terhadap variabel output. Berdasarkan seleksi variabel awal yang telah dilakukan, unsur A pada data set yang digunakan terdiri dari unsur Ti dan Zr, sedangkan unsur B terdiri dari unsur Mn, Co, Cr, V, Ni, dan Sn. Variabel dengan unsur-unsur yang dikelompokkan menjadi unsur A dan B ini tidak akan melalui proses feature engineering karena seluruh unsur A dan B tidak bernilai nol.

4.1.2. Preprocessing Data

Data *preprocessing* dilakukan untuk meningkatkan kualitas dataset yang dimiliki dengan cara mengatasi data yang masih hilang atau kekurangan atribut tertentu, data yang mengandung data outlier, dan mengatasi variasi nilai dataset yang terlalu besar. Kedua jenis variable akan mengalami preprocessing data yang sama. Dilakukan beberapa jenis *data preprocessing* pada penelitian ini yaitu, pengecekan *null*, *scaling*, pendeteksian dan penanganan *outliers*, serta pengujian multikolinearitas untuk variabel input.

a. Pengecekan *Null*

Pengecekan null dilakukan untuk mengetahui apakah ada nilai yang hilang pada setiap variabel. Nilai yang hilang ini dapat menyebabkan terjadinya *error* pada saat proses fitting data. Pengecekan *null* menggunakan fungsi "isnull" akan menghasilkan dua kemungkinan output yaitu *true* apabila terdapat nilai *null* pada variabel dan *false* apabila tidak terdapat nilai *null*. Hasil pengecekan didapatkan sebagai berikut:

```
#check for nulls
display("--*100)
display(AB2.isnull().any())
```

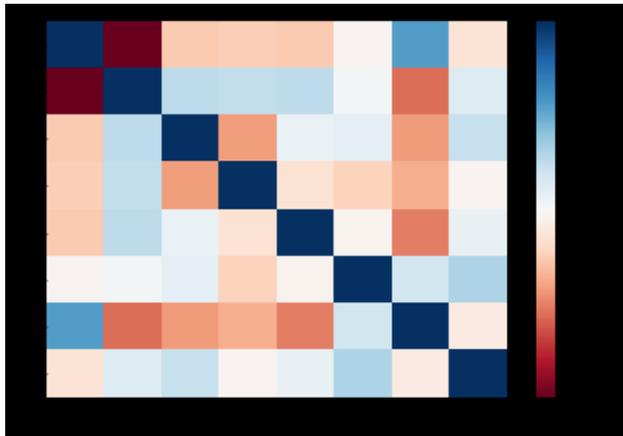
AB2

Ti	False
Zr	False
Mn	False
Co	False
Cr	False
V	False
Ni	False
Sn	False
ΔH abs	False
ΔS abs	False
C14	False
C15	False

Gambar 4.5 Pengecekan *Null*

Seluruh variabel menghasilkan output false yang berarti bahwa tidak terdapat nilai *null* pada variabel sehingga pengujian bisa dilanjutkan ke tahap selanjutnya.

b. Uji Multikolinearitas



Gambar 4.6 Heat Map Uji Kolerasi Variabel Input

Multikolinearitas terjadi ketika terdapat satu variabel input yang memiliki korelasi yang sangat tinggi dengan variabel inout yang lain. Hal ini dapat menyebabkan redundansi yang berujung

pada variabel yang tidak signifikan secara statistik. Untuk menguji apakah ada multikolinearitas, digunakan matriks korelasi, matriks korelasi antara setiap variabel input ini diletakkan pada sebuah heat map untuk mempermudah visualisasi data yang memiliki multikolinearitas. Warna yang gelap menunjukkan variabel yang memiliki korelasi yang tinggi satu dengan yang lain.

Pengujian korelasi yang dilakukan terhadap dataset dengan variabel unsur yang tidak dikelompokkan menghasilkan heatmap seperti gambar 4.6 di atas. Didapatkan bahwa unsur Ti dan Zr memiliki korelasi yang sangat tinggi satu dengan yang lain seperti yang dilingkari oleh lingkaran berwarna merah. Namun, variabel Ti dan Zr tidak akan dihapuskan dari model walaupun Ti dan Zr bersifat multikolinear satu dengan yang lain. Hal ini dapat terjadi karena unsur Ti dan Zr adalah anggota dari senyawa A pada AB_2 yang mana senyawa A akan selalu bernilai 1. Sehingga Ti bersifat berbanding terbalik dengan Zr, apabila komposisi Ti menurun berarti komposisi Zr menaik, dan sebaliknya. Sedangkan, variabel dengan unsur-unsur yang dikelompokkan menjadi A dan B tidak menunjukkan adanya multikolinearitas antara variabel A dan B.

c. Deteksi dan Penanganan Outliers

Outliers dapat memberikan efek yang tidak proporsional terhadap hasil statistik data seperti nilai mean yang dapat berujung kepada interpretasi model yang salah. Untuk mendeteksi adanya outliers, ditambahkan sebuah *metric* baru berupa 3 standar deviasi dibawah dan diatas nilai mean masing masing variabel. Selanjutnya akan seluruh dataset akan difilter berdasarkan nilai Z, yaitu jarak suatu data terhadap rata-ratanya dalam satuan standar deviasi. Apabila nilai Z diatas $+3$ standar deviasi, maka nilai tersebut akan dianggap sebagai outliers yang akan dihapus

```

# get the summary
desc_AB2 = AB2.describe()

#tambah metric standar deviation +-3
desc_AB2.loc['+3_std'] = desc_AB2.loc['mean'] + (desc_AB2.loc['std'] * 3)
desc_AB2.loc['-3_std'] = desc_AB2.loc['mean'] - (desc_AB2.loc['std'] * 3)

# display it
desc_AB2

#remove yg diatas +-3 sd
AB2_remove_df = AB2[(np.abs(stats.zscore(AB2)) < 3).all(axis=1)]

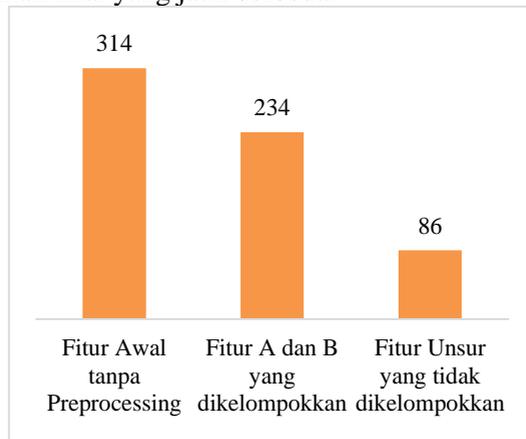
# what rows were removed
AB2.index.difference(AB2_remove_df.index)

Int64Index([30, 31, 32, 48, 49, 70, 71, 72, 75, 81, 86], dtype='int64')

```

Gambar 4.7 Deteksi dan Penanganan Outliers

Gambar 4.7 adalah contoh deteksi dan penanganan outlier yang dilakukan untuk variable dengan unsur-unsur yang tidak dikelompokkan. Pada proses tersebut, ditemukan 11 baris dataset yang mengandung outliers seperti yang dapat dilihat pada gambar bagian bawah yaitu baris ke-30, 31, 32, 48, 49, 70, 71, 72, 75, 81, 86, sehingga outliers ini dapat dihapus. Dapat dilihat bahwa baris outliers ini terletak berdekatan satu dengan yang lain seperti baris ke-30, 31, dan 32, hal ini kemungkinan besar terjadi karena data tersebut bersumber dari paper dan peneliti yang sama, namun set up pengujian yang ada pada paper tersebut berbeda dengan mayoritas set up pengujian yang ada pada dataset, sehingga menghasilkan nilai yang jauh berbeda.



Gambar 4.8 Jumlah Data Set untuk Setiap Variabel

Setelah dilakukan pendeteksian terhadap kedua jenis variabel, dapat dilihat hasil akhir jumlah dataset tanpa outliers pada gambar 4.7. Dari data yang awalnya berjumlah 314 pasang dataset, menjadi 234 pasang data untuk fitur A dan B yang dikelompokkan, dan 86 pasang data untuk fitur unsur yang tidak dikelompokkan. Dataset ini adalah dataset final yang akan digunakan sebagai input model machine learning.

d. Data Scaling

```
# DATA SCALING

scaler = preprocessing.StandardScaler().fit(AB2)
scaler_mean, scaler_scale = scaler.mean_, scaler.scale_
AB2_scaled = scaler.transform(AB2)
```

Gambar 4.9 Data Scaling

	x_train	x_train_minmax																																																																																																																																																									
	<table border="1"> <thead> <tr> <th></th> <th>Ti</th> <th>Zr</th> <th>Mn</th> <th>Co</th> <th>Cr</th> <th>V</th> <th>Ni</th> <th>Sn</th> </tr> </thead> <tbody> <tr><td>13</td><td>12.0</td><td>20.2</td><td>4.2</td><td>8.2</td><td>3.6</td><td>10.2</td><td>41.0</td><td>0.3</td></tr> <tr><td>1</td><td>12.9</td><td>21.7</td><td>5.5</td><td>1.5</td><td>8.4</td><td>9.8</td><td>39.5</td><td>0.3</td></tr> <tr><td>26</td><td>12.0</td><td>21.5</td><td>8.1</td><td>3.0</td><td>7.5</td><td>10.0</td><td>32.2</td><td>0.3</td></tr> <tr><td>3</td><td>12.0</td><td>20.2</td><td>5.7</td><td>1.5</td><td>9.7</td><td>10.2</td><td>41.0</td><td>0.3</td></tr> <tr><td>22</td><td>12.0</td><td>21.5</td><td>8.1</td><td>7.0</td><td>7.5</td><td>10.0</td><td>32.2</td><td>0.3</td></tr> <tr><td>8</td><td>12.0</td><td>20.2</td><td>5.2</td><td>5.1</td><td>5.6</td><td>10.2</td><td>41.0</td><td>0.3</td></tr> <tr><td>0</td><td>13.3</td><td>22.4</td><td>5.4</td><td>1.5</td><td>8.2</td><td>9.7</td><td>38.9</td><td>0.3</td></tr> <tr><td>14</td><td>11.7</td><td>19.6</td><td>4.2</td><td>8.3</td><td>3.6</td><td>10.3</td><td>41.6</td><td>0.3</td></tr> <tr><td>23</td><td>12.0</td><td>21.5</td><td>8.1</td><td>6.0</td><td>7.5</td><td>10.0</td><td>32.2</td><td>0.3</td></tr> <tr><td>33</td><td>12.0</td><td>21.5</td><td>13.6</td><td>0.5</td><td>4.5</td><td>10.0</td><td>37.6</td><td>0.3</td></tr> <tr><td>7</td><td>12.5</td><td>21.0</td><td>5.1</td><td>5.0</td><td>5.5</td><td>10.0</td><td>40.2</td><td>0.3</td></tr> <tr><td>11</td><td>12.9</td><td>21.7</td><td>4.0</td><td>7.9</td><td>3.4</td><td>9.8</td><td>39.5</td><td>0.3</td></tr> <tr><td>2</td><td>12.5</td><td>21.0</td><td>5.6</td><td>1.5</td><td>8.5</td><td>10.0</td><td>40.2</td><td>0.3</td></tr> <tr><td>9</td><td>11.7</td><td>19.6</td><td>5.3</td><td>5.2</td><td>5.7</td><td>10.3</td><td>41.6</td><td>0.3</td></tr> <tr><td>35</td><td>12.0</td><td>21.5</td><td>13.6</td><td>1.5</td><td>4.5</td><td>10.0</td><td>36.6</td><td>0.3</td></tr> <tr><td>15</td><td>12.0</td><td>21.5</td><td>8.1</td><td>8.0</td><td>7.5</td><td>10.0</td><td>32.2</td><td>0.3</td></tr> </tbody> </table>		Ti	Zr	Mn	Co	Cr	V	Ni	Sn	13	12.0	20.2	4.2	8.2	3.6	10.2	41.0	0.3	1	12.9	21.7	5.5	1.5	8.4	9.8	39.5	0.3	26	12.0	21.5	8.1	3.0	7.5	10.0	32.2	0.3	3	12.0	20.2	5.7	1.5	9.7	10.2	41.0	0.3	22	12.0	21.5	8.1	7.0	7.5	10.0	32.2	0.3	8	12.0	20.2	5.2	5.1	5.6	10.2	41.0	0.3	0	13.3	22.4	5.4	1.5	8.2	9.7	38.9	0.3	14	11.7	19.6	4.2	8.3	3.6	10.3	41.6	0.3	23	12.0	21.5	8.1	6.0	7.5	10.0	32.2	0.3	33	12.0	21.5	13.6	0.5	4.5	10.0	37.6	0.3	7	12.5	21.0	5.1	5.0	5.5	10.0	40.2	0.3	11	12.9	21.7	4.0	7.9	3.4	9.8	39.5	0.3	2	12.5	21.0	5.6	1.5	8.5	10.0	40.2	0.3	9	11.7	19.6	5.3	5.2	5.7	10.3	41.6	0.3	35	12.0	21.5	13.6	1.5	4.5	10.0	36.6	0.3	15	12.0	21.5	8.1	8.0	7.5	10.0	32.2	0.3	<pre>array([[0.1875, 0.21428571, 0.02083333, 0.98717949, 0.03174603, 0.83333333, 0.93617021, 0.], [0.75, 0.75, 0.15625, 0.12820513, 0.79365079, 0.16666667, 0.77659574, 0.], [0.1875, 0.67857143, 0.42708333, 0.32051282, 0.65079365, 0.5, 0.], [0.1875, 0.21428571, 0.17708333, 0.12820513, 1. , 0.83333333, 0.93617021, 0.], [0.1875, 0.67857143, 0.42708333, 0.83333333, 0.65079365, 0.5, 0.], [0.1875, 0.21428571, 0.125, 0.58974359, 0.34920635, 0.83333333, 0.93617021, 0.], [1. , 1. , 0.14583333, 0.12820513, 0.76190476, 0. , 0.71276596, 0.], [0. , 0. , 0.02083333, 1. , 0.03174603, 1. , 1. , 0.], [0.1875, 0.67857143, 0.42708333, 0.70512821, 0.65079365, 0.5, 0.], [0.1875, 0.67857143, 1. , 0. , 0.17460317, 0.5, 0.57446809, 0.], [0.5, 0.5, 0.11458333, 0.57692308, 0.33333333, 0.5, 0.85106383, 0.], [0.75, 0.75, 0. , 0.94871795, 0. , 0.16666667, 0.77659574, 0.], [0.5, 0.5, 0.16666667, 0.12820513, 0.80952381, 0.5, 0.85106383, 0.], [0. , 0. , 0.13541667, 0.6025641, 0.36507937,</pre>
	Ti	Zr	Mn	Co	Cr	V	Ni	Sn																																																																																																																																																			
13	12.0	20.2	4.2	8.2	3.6	10.2	41.0	0.3																																																																																																																																																			
1	12.9	21.7	5.5	1.5	8.4	9.8	39.5	0.3																																																																																																																																																			
26	12.0	21.5	8.1	3.0	7.5	10.0	32.2	0.3																																																																																																																																																			
3	12.0	20.2	5.7	1.5	9.7	10.2	41.0	0.3																																																																																																																																																			
22	12.0	21.5	8.1	7.0	7.5	10.0	32.2	0.3																																																																																																																																																			
8	12.0	20.2	5.2	5.1	5.6	10.2	41.0	0.3																																																																																																																																																			
0	13.3	22.4	5.4	1.5	8.2	9.7	38.9	0.3																																																																																																																																																			
14	11.7	19.6	4.2	8.3	3.6	10.3	41.6	0.3																																																																																																																																																			
23	12.0	21.5	8.1	6.0	7.5	10.0	32.2	0.3																																																																																																																																																			
33	12.0	21.5	13.6	0.5	4.5	10.0	37.6	0.3																																																																																																																																																			
7	12.5	21.0	5.1	5.0	5.5	10.0	40.2	0.3																																																																																																																																																			
11	12.9	21.7	4.0	7.9	3.4	9.8	39.5	0.3																																																																																																																																																			
2	12.5	21.0	5.6	1.5	8.5	10.0	40.2	0.3																																																																																																																																																			
9	11.7	19.6	5.3	5.2	5.7	10.3	41.6	0.3																																																																																																																																																			
35	12.0	21.5	13.6	1.5	4.5	10.0	36.6	0.3																																																																																																																																																			
15	12.0	21.5	8.1	8.0	7.5	10.0	32.2	0.3																																																																																																																																																			

Gambar 4.10 Perbandingan Data (a) sebelum scaling dan (b) setelah scaling

Data scaling adalah suatu cara untuk membuat numerical data pada dataset memiliki rentang nilai (*scale*) yang sama, sehingga tidak ada lagi satu variabel data yang mendominasi variabel data lainnya. Pada penelitian ini digunakan proses data scaling terhadap nilai mean yang ada menggunakan fungsi scaler.mean . Perbandingan data sebelum dan sesudah scaling dapat dilihat pada gambar 4.10a. Dimana dapat diamati sebelum dilakukannya scaling masih terdapat perbedaan nilai yang sangat besar antara variabel input yang ada seperti unsur Nikel yang dapat mencapai harga 41. Scaling tersebut dilakukan menggunakan rumus berikut:

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \dots \dots \dots (10)$$

dimana X' adalah nilai fitur yang telah mengalami scaling, X adalah nilai fitur sebelum scaling, X_{\min} adalah nilai minimum dari dataset fitur lama dan X_{\max} adalah nilai maksimum dari dataset fitur lama. Sebagai contoh, dilakukan perhitungan untuk melakukan scaling pada salah satu unsur Mangan yang bernilai 4,2. Diketahui bahwa nilai mangan minimum pada dataset adalah, dan nilai mangan maksimum pada dataset adalah 13,6 dan nilai minimum mangan pada dataset adalah 4 maka,

$$\begin{aligned} X' &= \frac{X - X_{\min}}{X_{\max} - X_{\min}} \\ X' &= \frac{4,2 - 4}{13,6 - 4} \\ X' &= 0,02083 \end{aligned}$$

setelah dilakukannya scaling ditemukan nilai mangan yang telah mengalami scaling sebesar 0,02083. Rumus ini akan membuat semua data berada pada rentang 0-1 seperti yang dapat dilihat pada gambar 4.10b dimana hal ini akan memberikan kesempatan yang sama bagi seluruh variabel input untuk berkontribusi terhadap model yang ada.

e. Pemilihan Jumlah Tree untuk Model Random Forests

Dilakukan pengujian untuk menentukan jumlah tree yang dapat memberikan prediksi yang paling optimum bagi setiap variabel output untuk model random forest. Pengujian dilakukan dengan melakukan pengecekan prediksi untuk model dengan jumlah tree mulai dari 0 sampai dengan 100. Hasil dari pengujian diletakkan pada grafik dimana sumbu x menunjukkan jumlah tree, dan sumbu y menunjukkan kemampuan prediksi. Jumlah tree yang dipilih adalah yang memiliki kemampuan prediksi terbaik, ditandai dengan titik dengan posisi yang paling tinggi pada grafik. Pengujian untuk ΔH absorpsi, ΔS absorpsi, phase abundance, dan persen massa hidrogen menemukan jumlah tree optimal adalah 21, 8, 32, dan 7 tree secara berurutan. Jumlah tree ini dipakai dalam penentuan jumlah tree pada saat pembuatan model random forest.

4.3. Modelling Variabel dengan Unsur-Unsur yang Tidak Dikelompokkan

Didapatkan hasil evaluasi dari masing masing model yaitu regresi linear multivariate, decision tree, dan random forest untuk variabel dengan unsur-unsur yang tidak dipisahkan. Hasil dari setiap model dibandingkan berdasarkan setiap variable output guna menemukan algoritma terbaik untuk memprediksi setiap variable output. Pengaruh masing masing unsur terhadap setiap variabel output didapatkan untuk mengetahui unsur-unsur yang dapat memberikan keuntungan bagi sifat-sifat penyimpanan hydrogen logam hidrida.

4.3.1. Perbandingan Hasil Evaluasi Algoritma untuk Prediksi Selisih Entalpi

Tabel 4.2 Perbandingan Metode Machine Learning untuk Prediksi Selisih Entalpi Pembentukan

Metric	Multivariate Regression	Decision Tree	Random Forest
---------------	--------------------------------	----------------------	----------------------

R ² Train	0.478	0.868	0.819
R ² Validasi	0.055	-0.05	0.591
R ² Test	0.569	0.22	0.531
R ² Rata-rata	0.367	0.346	0.647
Mean Absolute Error	4,61	5,9	4,36
Root Mean Square Error	5.51	7,39	5,75



Gambar 4.11 Grafik Nilai Asli dan Nilai Prediksi untuk Selisih Entalphi Pembentukan

Ketiga model machine learning dibandingkan untuk melihat performansi pada prediksi selisih entalphi pembentukan.

Algoritma multivariate regression menghasilkan fungsi matematis sebagai berikut,

$$Y = 28,51 + 19,82Ti + 8,68Zr + 2,22Mn - 6,51Co + 4,54Cr - 4,54V - 15,15Ni + 5,92Sn \dots \dots \dots (11)$$

dimana jumlah mol masing masing unsur dapat dimasukkan kedalam fungsi matematis tersebut untuk menemukan nilai Y yang merupakan prediksi delta H dari data yang baru. Model random forest dan decision tree juga menghasilkan suatu fungsi matematis yang dapat digunakan untuk memprediksi dengan memasukkan data input yang baru ke software python. Untuk melakukan evaluasi terhadap model yang didapatkan, dilakukan pengecekan terhadap R^2 dan nilai error yang dimiliki model. R^2 rata-rata atau kemampuan prediksi tertinggi diraih oleh random forest pada angka 0,647 diikuti oleh multivariate linear regression pada 0.367 dan decision tree pada angka 0.346. Hal ini menandakan bahwa random forest adalah model yang paling baik digunakan untuk melakukan prediksi selisih entalphi pembentukan, serta dataset yang dimiliki bersifat non-linear sehingga apabila ingin dilakukan penelitian lebih lanjut akan lebih baik untuk menggunakan algoritma non-linear. Walaupun telah dilakukan preprocessing data dan feature selection, nilai R^2 rata-rata model masih belum bisa menyentuh angka 0,70. Hal ini menandakan bahwa kemampuan prediksi delta entalphi masih tergolong lemah. Dapat dilihat bahwa terdapat penurunan nilai R^2 pada data validasi apabila dibandingkan dengan data training, hal ini sebenarnya lumrah terjadi pada metode analisa machine learning, namun penurunan skor R^2 yang sangat besar menandakan terjadinya overfitting yang sangat besar pula pada model tersebut. Pada prediksi delta entalphi ini dapat diamati bahwa terjadi overfitting yang besar pada model multivariate regression dan decision tree.

Mean absolute error tertinggi ada pada decision tree pada 5,9 diikuti multivariate regression pada angka 4,61 dan random

forest pada 4,36. Root mean square error tertinggi ada pada decision tree pada angka 7,39, diikuti oleh random forest pada angka 5,75 dan multivariate regression pada angka 5,51. Error yang terjadi terlihat besar karena error berada cukup jauh dari angka 0. Namun, apabila grafik plot nilai asli dan nilai prediksi dicermati, dapat dilihat bahwa selisih nilai yang didapatkan oleh prediksi model tidak terlalu jauh dengan nilai asli. Hasil prediksi model yang digambarkan oleh garis hijau, kuning, dan merah untuk linear regression, decision tree, dan random forest secara berurutan sudah dapat mengikuti fluktuasi grafik nilai asli yang digambarkan oleh garisbiru. Namun, dapat dilihat bahwa ketiga model gagal untuk memprediksi nilai selisih entalpi yang bernilai tinggi seperti yang berada pada angka 47 KJ/mol. Di sisi lain, model decision tree dan random forest bekerja dengan baik dalam hal memprediksi nilai selisih entalpi yang kecil seperti yang bernilai dibawah 10 KJ/mol. Tentu saja, berdasarkan model yang didapatkan dapat diketahui bahwa akan ada kemungkinan terjadi kesalahan pada saat melakukan prediksi yang baru. Namun, walaupun pada akhirnya terjadi selisih yang sangat besar antara nilai prediksi dan nilai asli selisih entalpi pembentukan. Konsekuensi yang dihasilkan dari error ini tidak akan terlalu besar, karena model prediksi hanya akan digunakan untuk tujuan penelitian, sehingga tidak akan membahayakan bagi manusia. Oleh karena itu nilai error dan R^2 yang cukup lemah ini masih dapat diterima. Selisih entalpi pembentukan untuk alloy yang akan diteliti masih akan ditemukan dan tetap akan menjadi informasi yang berguna bagi peneliti yang lainnya sebagai baseline sebelum melakukan eksperimen pemaduan logam. Selain itu apabila selisih entalpi yang diinginkan justru terlalu besar, hal ini masih dapat dikompensasi dengan properti hidrogenasi yang lain seperti persen massa hidrogen yang dihasilkan, karena kemungkinan apabila energi pembentukan yang dibutuhkan semakin besar, maka persen massa hidrogen yang dihasilkan akan semakin besar pula yang merupakan hal yang diinginkan pada logam hidrida.

4.3.2. Perbandingan Hasil Evaluasi Algoritma untuk Prediksi Selisih Entropi

Tabel 4.3 Perbandingan Metode Machine Learning untuk Prediksi Selisih Entropi Pembentukan

Metric	Multivariate Regression	Decision Tree	Random Forest
R^2 Train	0.332	0.935	0.868
R^2 Validasi	-0.27	-1.5	0.6922
R^2 Test	0.469	0.06	0.104
R^2 Rata-rata	0.177	-0.16	0.554
Mean Absolute Error	9.39	7,57	30,8
Root Mean Square Error	13.0	14,2	33,6



Gambar 4.12 Grafik Nilai Asli dan Nilai Prediksi untuk Selisih Entropi Pembentukan

Ketiga model machine learning dibandingkan untuk melihat performansinya pada prediksi selisih entropi pembentukan. Algoritma multivariate regression menghasilkan fungsi matematis sebagai berikut,

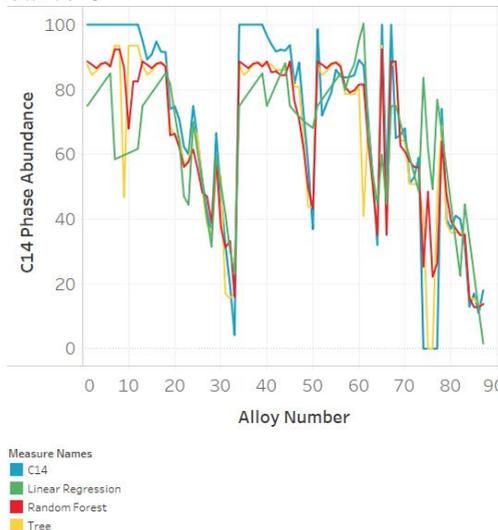
$$Y = 89,49 + 44,62Ti + 44,88Zr + 0,54Mn - 15,55Co + 2,82Cr - 7,08V - 31,21Ni - 0,54Sn.....(12)$$

dimana jumlah mol masing masing unsur dapat dimasukkan kedalam fungsi matematis tersebut untuk menemukan nilai Y yang merupakan prediksi delta S dari data yang baru, hal yang sama juga berlaku untuk model decision tree dan random forest, dimana nilai input yang baru dapat dimasukkan ke software python untuk menemukan prediksi delta entropi yang baru. Nilai R^2 dan error dilihat untuk mengevaluasi model yang didapatkan. Nilai R^2 rata-rata tertinggi diraih oleh model random forest pada angka 0,554 diikuti oleh multivariate regression pada angka 0,177, dan decision tree pada angka 0,16. Hal ini menandakan bahwa random forest adalah model yang paling baik untuk memprediksi selisih entropi, serta dataset yang dimiliki bersifat non-linear. Ketiga model yang didapatkan menghasilkan R^2 rata-rata yang bernilai dibawah 0,7 yang menandakan bahwa prediksi yang dihasilkan masih cukup lemah. Terjadi overfitting yang besar pada model multivariate regression dan decision tree yang ditandai dengan penurunan nilai R^2 validasi yang sangat besar apabila dibandingkan dengan R^2 train.

Nilai error yang didapat tergolong cukup besar pada ketiga model, dimana untuk mean absolute error didapatkan nilai sebesar 9,39, 7,57, dan 30,8 untuk multivariate regression, decision tree, dan random forest secara berurutan. Root mean square error menunjukkan nilai 13,0, 14,2, dan 33,6 untuk multivariate regression, decision tree, dan random forest secara berurutan. Error yang terjadi terlihat jauh lebih besar apabila dibandingkan dengan error pada prediksi selisih entalpi pembentukan. Hal ini dapat

terjadi karena memang pada proses pengumpulan data, banyak dataset dengan nilai selisih entropi pembentukan yang diasumsikan sebagai 110 KJ/mol untuk mendapatkan nilai selisih entalpi pembentukan. Nilai selisih entropi pembentukan dipilih sebagai nilai yang diasumsikan karena memang sebenarnya nilai entropi tidak berkaitan secara langsung dengan energi pembentukan. Nilai selisih entropi dihasilkan akibat terjadinya perubahan wujud hidrogen dari yang semula berbentuk gas menjadi padat ketika diserap oleh logam hidrida. Sehingga error yang cukup besar ini dapat ditoleransi. Apabila grafik plot nilai asli dan nilai prediksi dilihat, model sebenarnya telah dapat mengikuti fluktuasi nilai asli. Terutama model decision tree dan random forest yang bekerja sangat bagus dalam memprediksi nilai selisih entropi pada 110 KJ/mol yang sebenarnya diasumsikan oleh peneliti.

4.3.3. Perbandingan Hasil Evaluasi Algoritma untuk Prediksi Phase Abundance C14



Gambar 4.13 Grafik Nilai Asli dan Nilai Prediksi untuk Phase Abundance C14

Tabel 4.4 Perbandingan Metode Machine Learning untuk Prediksi Phase Abundance C14

Metric	Multivariate Regression	Decision Tree	Random Forest
R ² Train	0.419	0.966	0.893
R ² Validasi	0.292	0.788	0.819
R ² Test	0.633	0.603	0.786
R ² Rata-rata	0.448	0.785	0.832
Mean Absolute Error	11.5	8.48	22,4
Root Mean Square Error	13,3	13.8	25,6

Prediksi model yang dilakukan terhadap variabel output C14 menghasilkan R² rata-rata tertinggi pada random forest di angka 0.832, diikuti oleh decision tree pada angka 0,785, dan multivariate regression pada angka 0,448. Model decision tree dan random forest memiliki kemampuan diatas 0,70 yang berarti bahwa model tersebut memiliki kualitas yang baik untuk melakukan prediksi phase abundance C14. Apabila nilai R² train dan R² validasi dibandingkan, dapat dilihat bahwa penurunan nilai R² pada data validasi tidak terlalu besar untuk ketiga model, hal ini menandakan bahwa overfitting yang terjadi tidak terlalu besar yang menandakan performansi yang baik bagi model. Algoritma multivariate regression meghasilkan fungsi matematis sebagai berikut,

$$Y = 32,47 + 17,66\text{Ti} + 14,18\text{Zr} - 19,25\text{Mn} - 24,99\text{Co} + 2,34\text{Cr} + 44,44\text{V} - 54,67\text{Ni} + 61,86\text{Sn} \dots\dots\dots(13)$$

dimana jumlah mol masing masing unsur dapat dimasukkan kedalam fungsi matematis tersebut untuk menemukan nilai Y yang merupakan prediksi phase abundance C14 dari data yang baru.

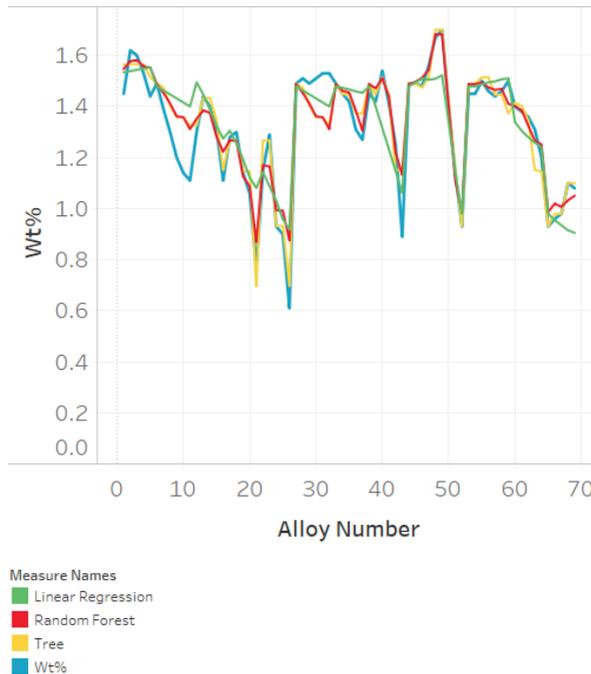
Nilai mean absolute error tertinggi diraih oleh model random forest pada angka 22,4, diikuti oleh multivariate regression pada angka 11,5 dan decision tree pada angka 8,48. Nilai root mean square error tertinggi juga diraih oleh random forest pada angka 25,6, diikuti decision tree pada angka 13,8, dan multivariate regression pada angka 13,3. Nilai error baik yang dihasilkan oleh mean absolute error maupun root mean square error tergolong besar karena sangat jauh dari nilai 0. Performa prediksi digrafikan pada grafik nilai asli dan nilai prediksi, dimana dapat dilihat bahwa sebenarnya model sudah cukup baik dalam mengikuti naik turunnya nilai asli. Decision tree memiliki performansi yang sangat baik untuk memprediksi nilai C14 yang rendah, dimana model decision tree adalah satu satunya model yang dapat meramalkan phase abundance C14 yang bernilai 0. Nilai error yang besar yang dihasilkan pada prediksi ini mungkin terjadi karena banyak logam paduan dari dataset yang memiliki phase C14 sebanyak 100% dan 0%, tentu saja model machine learning akan lebih susah memprediksi nilai yang sangat tinggi maupun sangat rendah, karena nilai tersebut terletak sangat jauh dari nilai median dan dapat terlihat seperti outliers walaupun pada nyatanya memang benar bahwa suatu logam paduan dapat menghasilkan nilai 100% maupun 0% phase abundance C14.

4.3.4. Perbandingan Hasil Evaluasi Algoritma untuk Prediksi Persen Massa Hidrogen

Tabel 4.5 Perbandingan Metode Machine Learning untuk Prediksi Persen Massa Hidrogen

Metric	Multivariate Regression	Decision Tree	Random Forest
R ² Train	0.741	0.989	0.928
R ² Validasi	0.608	0.113	0.521
R ² Test	0.654	0.392	0.615
R ² Rata-rata	0.667	0.498	0.688

Mean Absolute Error	0,0968	0,0894	0,101
Root Mean Square Error	0,12	0,159	0,126



Gambar 4.14 Grafik Nilai Asli dan Nilai Prediksi untuk Persen Massa Hidrogen

Persen massa hidrogen adalah salah satu sifat yang paling penting dalam aplikasi logam hidrida, karena persen massa hidrogen akan mempengaruhi massa logam hidrida secara keseluruhan. Pada prediksi machine learning yang dilakukan

terhadap persen massa hidrogen, random forest meraih nilai R^2 rata-rata tertinggi pada angka 0,688, diikuti oleh multivariate regression pada angka 0,667, dan decision tree pada angka 0,498. Ketiga model tidak menyentuh angka 0,70 yang berarti bahwa kemampuan prediksi yang dimiliki masih tergolong lemah. Overfitting yang sangat besar dapat diamati pada model decision tree dengan penurunan skor R^2 validasi yang cukup signifikan apabila dibandingkan dengan nilai R^2 training. Algoritma multivariate regression menghasilkan fungsi matematis sebagai berikut,

$$Y = 0,31 + 0,38Ti - 0,06Zr + 0,72Mn + 0,18Co + 0,42Cr - 0,05V + 0,12Ni + 0,29Sn \dots \dots \dots (14)$$

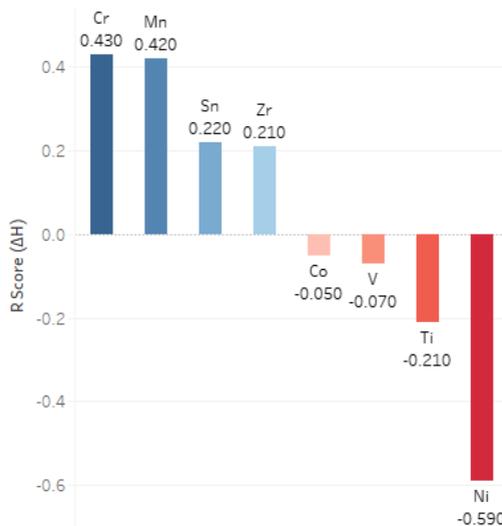
dimana jumlah mol masing masing unsur dapat dimasukkan kedalam fungsi matematis tersebut untuk menemukan nilai Y yang merupakan prediksi nilai persen massa hidrogen dari data yang baru. Mean absolute error tertinggi diraih oleh random forest pada angka 0,101, diikuti oleh multivariate regression pada angka 0,0968, dan decision tree pada angka 0,0894.. Sedangkan root mean square tertinggi diraih oleh decision tree pada 0,159, random forest pada 0,126 dan multivariate regression pada 0,126. Ketiga metode pengujian error ini memberikan hasil yang bagus karena error berada pada angka yang mendekati nilai 0. Nilai error yang kecil ini dapat terjadi karena berbeda dengan variabel entalphi dan entropi pembentukan dimana terdapat banyak nilai yang diasumsikan sehingga dapat mempengaruhi performa model karena nilai yang diasumsikan tersebut dapat mempengaruhi mean yang dimiliki oleh dataset secara keseluruhan, dan pada variabel phase abundance banyaknya dataset yang memiliki phase 100% dan 0% juga mempengaruhi model prediksi karena nilai tersebut terletak jauh dari mean yang dimiliki oleh dataset sehingga membuat model kesulitan untuk meraih nilai tersebut.

4.3.5. Pengaruh Unsur terhadap Heat of Formation

Pengaruh komposisi kimia terhadap heat of formation diteliti menggunakan klasifikasi regresi linear dan klasifikasi decision tree. Selisih entalpi dan selisih entropi pembentukan digunakan sebagai heat of formation yang diteliti.

4.3.5.1. Selisih Entalpi Pembentukan

a. Klasifikasi Berdasarkan Regresi Linear



Gambar 4.15 Grafik Pengaruh Elemen untuk ΔH

Klasifikasi regresi linear menghasilkan data yang tertera pada tabel 4.2 Dan pada tabel 4.3 Bagi data yang sudah diurutkan. Dari tabel *ranking* dapat dilihat bahwa unsur Ni memberikan efek yang paling baik bagi nilai selisih entalpi pembentukan dimana penambahan Ni dapat menurunkan nilai tersebut. *Ranking* selanjutnya diduduki oleh Ti, V, Co, Zr, Sn, Mn, dan Cr secara berurutan. Sesuai dengan teori, penambahan Ni akan menurunkan

energi pembentukan logam hidrida. Hal ini disebabkan karena nikel merupakan unsur yang pada umumnya digunakan sebagai katalis dalam proses hidrogenasi. Penambahan nikel pada logam paduan AB_2 berbasis titanium akan menyebabkan terjadinya ikatan yang lebih kuat antara unsur titanium dan unsur-unsur penyusun B, ikatan yang lebih kuat ini akan menyebabkan menurunnya muatan transfer antara logam dengan hidrogen. Hal ini akan berkontribusi terhadap menurunnya tekanan pembentukan logam hidrida. Tanpa adanya katalis, reaksi pembentukan logam hidrida akan berlangsung pada temperatur dan tekanan yang sangat tinggi, penambahan katalis akan mengurangi tekanan dan temperatur pembentukan yang berarti bahwa energi yang dibutuhkan untuk proses pembentukan akan turut berkurang, oleh karena itulah nikel pada saat ini digunakan sebagai unsur paduan pada baterai hidrogen komersial berupa Nickel Metal Hydride atau Ni-MH. Posisi terakhir diduduki oleh kromium dimana klasifikasi regresi linear menunjukkan kenaikan nilai energi pembentukan, hal ini didukung oleh hasil percobaan yang dilakukan oleh Kim et. al [18] bahwa pada sistem paduan Zr-Cr-Ni-La, penambahan kromium dapat membentuk lapisan oksida yang cukup tebal sehingga dapat menghambat penetrasi hidrogen yang berujung pada periode aktivasi atau hidrogenasi yang lebih lama serta energi pembentukan yang lebih tinggi.

b. Klasifikasi Berdasarkan Feature Importance Decision Tree

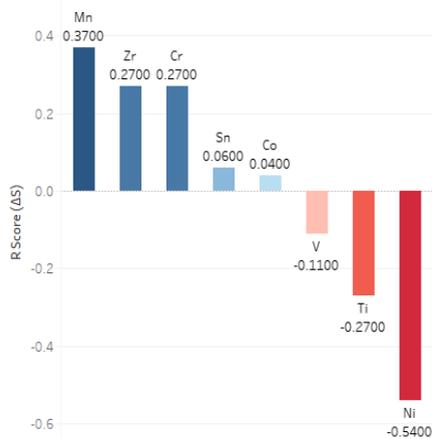
Tabel 4.6 Terjemahan Decision Tree untuk Selisih Entalpi Absorpsi

Rank	Unsur	Batas 1	Efek 1	Batas 2	Efek 2
1	Ni	$>1,911$	Menurunkan	$\leq 1,911$	Menaikkan
2	Cr	$\leq 0,217$	Menurunkan	$>0,217$	Menaikkan
2	Ti	$\leq 0,372$	Menaikkan	$>0,372$	Menurunkan
3	Co	$\leq 0,06$	Menaikkan	$>0,06$	Menurunkan
3	Mn	$\leq 0,243$	Menurunkan	$>0,243$	Menaikkan
4	V	$\leq 0,29$	Menaikkan	$>0,29$	Menurunkan
4	Sn	$\leq 0,009$	Menurunkan	$>0,009$	Menaikkan

Berdasarkan klasifikasi decision tree, dapat dilihat bahwa unsur nikel menduduki urutan pertama sebagai fitur yang dianggap paling penting untuk menentukan nilai selisih entalpi, diikuti oleh kromium dan titanium pada posisi kedua, kobalt dan mangan pada posisi ketiga, serta vanadium dan timah pada posisi keempat. Unsur zirkonium tidak muncul pada decision tree yang berarti bahwa unsur zirkonium tidak memiliki peranan dalam menentukan besarnya selisih nilai entalpi. Dari klasifikasi decision tree kita dapat melihat bahwa ternyata terdapat nilai tertentu bagi unsur Ni yang dapat memberikan pengaruh baik dalam menentukan nilai energi pembentukan logam hidrida. Unsur Ni yang bernilai di atas 1,911 mol akan memberikan pengaruh positif untuk menurunkan energi pembentukan, sedangkan ketika Ni bernilai dibawah 1,911 akan terjadi kenaikan terhadap nilai energi pembentukan. Unsur kromium berada pada posisi kedua dengan efek menaikkan energi pembentukan logam hidrida untuk kandungan di atas 0,217 mol. Hal ini sesuai dengan prediksi regresi linear bahwa unsur kromium memberikan pengaruh terburuk pada urutan ke 8. Hal ini dapat menjadi petunjuk bagus bagi peneliti yang akan datang ketika menentukan jumlah komposisi kimia bagi logam paduan AB_2 yang akan diteliti.

4.3.5.2. Selisih Entropi Pembentukan

a. Klasifikasi Berdasarkan Regresi Linear



Gambar 4.16 Grafik Pengaruh Elemen untuk ΔS

Sama halnya dengan klasifikasi terhadap selisih entalpi absorpsi, unsur nikel menduduki urutan pertama dalam hal memberikan efek yang paling positif terhadap penurunan nilai entropi pembentukan. Nilai entropi memiliki kaitan erat dengan perubahan wujud hidrogen dari gas ke padat pada saat proses pembentukan logam hidrida. Penambahan katalis akan mempercepat proses perubahan wujud hidrogen sehingga energi yang dibutuhkan akan berkurang karena proses yang lebih cepat. Berbeda dengan klasifikasi pengaruh komposisi terhadap selisih entalpi, pada klasifikasi pengaruh komposisi terhadap selisih entropi unsur mangan menempati urutan yang terburuk dalam hal penambahan unsur mangan akan menaikkan nilai entropi pembentukan. Pada saat proses pembentukan logam hidrida, energi digunakan untuk menyatukan unsur-unsur paduan yang ada, unsur-unsur paduan ini akan lebih mudah menyatu ketika memiliki gaya tarik menarik yang lebih besar, ketika unsur paduan lebih mudah untuk menyatu, maka energi yang digunakan untuk pembentukan tersebut akan lebih kecil. Dalam kasus interaksi antara unsur hidrogen dan mangan, gaya tarik menarik antara kedua unsur ini

sangatlah kecil, oleh karena itulah energi pembentukan yang dibutuhkan menjadi lebih besar.

b. Klasifikasi Berdasarkan Feature Importance Decision Tree

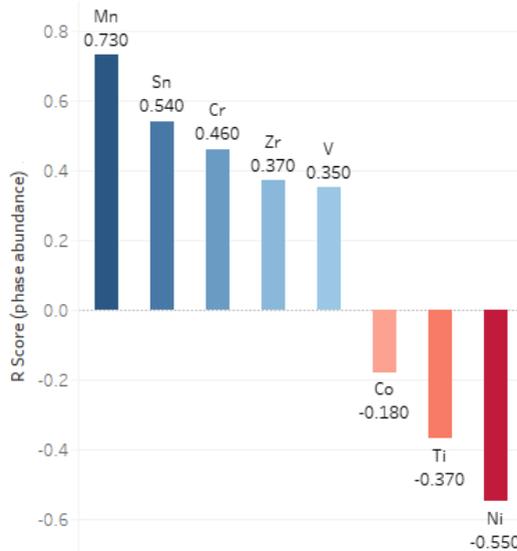
Tabel 4.7 Terjemahan Decision Tree untuk Selisih Entropi Absorpsi

Rank	Unsur	Batas 1	Efek 1	Batas 2	Efek 2
1	Ni	>1,911	Menurunkan	≤1,911	Menaikkan
2	Mn	≤0,409	Menaikkan	>0,409	Menurunkan
2	Ti	≤0,358	Menaikkan	>0,358	Menurunkan
3	V	≤0,298	Menaikkan	>0,298	Menurunkan
3	Sn	≤0,009	Menurunkan	>0,009	Menaikkan
4	Co	≤0,15	Menaikkan	>0,15	Menurunkan

Klasifikasi decision tree memberikan hasil yang tidak jauh berbeda dengan klasifikasi linear regression dimana nikel menduduki urutan pertama sebagai unsur yang dapat menurunkan nilai entropi, sedangkan mangan merupakan unsur yang dapat menaikkan nilai entropi. Penambahan unsur nikel yang dapat menurunkan nilai entropi adalah penambahan nikel diatas 1,911 mol. Sedangkan, penambahan unsur mangan yang dapat menaikkan nilai entropi adalah penambahan mangan diatas 0,409 mol. Klasifikasi decision tree tidak mengeluarkan hasil untuk unsur zirkonium dan kromium, yang berarti bahwa zirkonium dan kromium tidak memberikan pengaruh yang berarti terhadap nilai entropi pembentukan logam hidrida.

4.3.6. Pengaruh Unsur terhadap Phase Abundance C14

a. Klasifikasi Berdasarkan Regresi Linear



Gambar 4.17 Grafik Pengaruh Elemen untuk Phase Abundance

Klasifikasi regresi linear menghasilkan Mn sebagai unsur yang berada pada urutan pertama dalam hal menaikkan persen C14 pada phase abundance diikuti oleh Sn, Cr, Zr, V, Co, Ti, dan Ni secara berurutan. Mangan menempati posisi pertama dengan skor korelasi yang cukup tinggi yaitu sebesar 0,73. Hal ini sebenarnya tidak sesuai dengan teori karena unsur mangan seharusnya bersifat menurunkan fasa C14 dan menaikkan fasa C15, seperti penelitian yang dilakukan oleh K. Young et. al [19] untuk membandingkan karakteristik logam paduan yang didominasi fasa C14 dan C15 menunjukkan bahwa penambahan jumlah unsur mangan pada logam paduan justru akan menaikkan kadar C15, hal ini disebabkan karena unsur mangan dapat memberikan permukaan reaktif yang lebih besar yang mendukung terbentuknya fasa C15. Sedangkan

unsur nikel berada pada posisi terakhir yang berarti bahwa unsur nikel akan lebih berkontribusi terhadap pembentukan fasa C15. Logam paduan yang didominasi oleh fasa C15 diketahui dapat melepaskan hidrogen dengan lebih mudah pada proses discharge, hal ini cukup sesuai dengan teori dimana nikel bersifat sebagai katalis dalam hal pada proses charging dan discharging yang dapat berkontribusi terhadap lebih mudahnya hidrogen untuk lepas menjadi gas pada saat discharging. Percobaan yang dilakukan oleh K. Young et. al [19] juga mendukung hal tersebut dimana logam paduan yang ditunjukkan untuk didominasi fasa C15 memiliki kadar nikel yang lebih tinggi pada angka 38 mol dibandingkan dengan logam paduan C14 dengan kadar nikel pada angka 32 mol. Hal ini menunjukkan bahwa terdapat ketidaksesuaian dengan teori pada hasil klasifikasi menggunakan regresi linear, hal ini mungkin terjadi karena kemungkinan terjadinya under fitting pada saat regresi linear cukup tinggi karena persebaran data yang cukup luas sehingga model tidak dapat memberikan prediksi dengan akurat.

b. Klasifikasi Berdasarkan Feature Importance Decision Tree

Tabel 4.8 Terjemahan Decision Tree untuk Phase Abundance C14

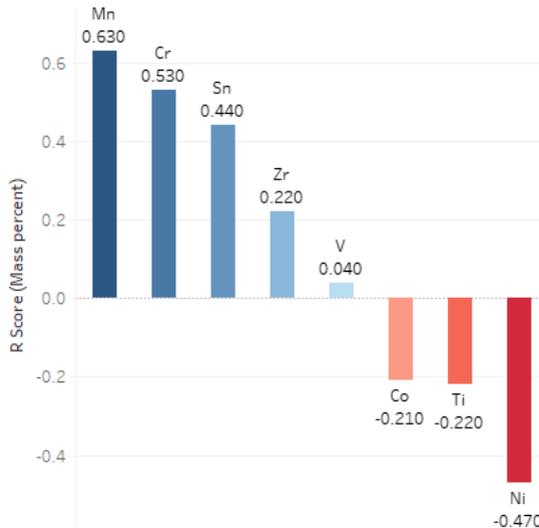
Rank	Unsur	Batas 1	Efek 1	Batas 2	Efek 2
1	Cr	$>0,135$	Menaikkan	$\leq 0,135$	Menurunkan
2	Ni	$\leq 0,120$	Menurunkan	$>0,120$	Menaikkan
2	Mn	$\leq 0,235$	Menurunkan	$>0,235$	Menaikkan
3	V	$\leq 0,292$	Menurunkan	$>0,292$	Menaikkan
3	Sn	$\leq 0,008$	Menurunkan	$>0,008$	Menaikkan
3	Ti	$\leq 0,372$	Menaikkan	$>0,372$	Menurunkan
4	Zr	$\leq 0,627$	Menaikkan	$>0,627$	Menurunkan
4	Co	$\leq 0,150$	Menaikkan	$>0,150$	Menurunkan

Berbeda dengan regresi linear, klasifikasi berdasarkan decision tree menunjukkan bahwa unsur kromium merupakan unsur yang paling berpengaruh dalam kenaikan persen fasa C14, sedangkan unsur Ti adalah unsur yang paling berpengaruh dalam menurunkan persen fasa C14. Posisi kromium sebagai unsur yang

paling berpengaruh sudah cukup sesuai dengan teori. Pada percobaan yang dilakukan oleh K. Young et.al [19], kromium digunakan sebagai unsur paduan utama bagi logam paduan dengan C14 sebagai fasa utama. Hal ini dilakukan karena unsur kromium merupakan bagian dari logam transisi dengan jumlah elektron valensi yang rendah. Hal ini akan menyebabkan menurunnya nilai rasio elektron per atom (e/a) yang berkontribusi kepada pembentukan fasa C14. Batas nilai e/a untuk menentukan fasa yang terbentuk adalah $e/a \sim 6,9$. Dibawah nilai 6,9 logam paduan akan membentuk lebih banyak fasa C14 dan diatas 6,9 logam paduan akan lebih cenderung untuk membentuk fasa C15. Terdapat perbedaan pada efek yang diberikan oleh unsur nikel berdasarkan klasifikasi regresi linear dengan klasifikasi decision tree. Pada klasifikasi decision tree nikel berfungsi sebagai unsur yang dapat menaikkan kadar fase C14. Perbedaan ini mungkin terjadi karena memang sebenarnya fungsi nikel adalah sebagai katalis, sedangkan yang paling mempengaruhi pembentukan C14/C15 adalah nilai rasio elektron per atom dari logam paduan. Unsur nikel sendiri tidak memberi pengaruh yang terlalu signifikan terhadap perubahan nilai e/a pada logam paduan.

4.3.7. Pengaruh Unsur terhadap Persen Massa Hidrogen

a. Klasifikasi Berdasarkan Regresi Linear



Gambar 4.18 Grafik Pengaruh Elemen untuk Kapasitas Hidrogen

Klasifikasi regresi linear untuk persen massa hidrogen menghasilkan Mn pada urutan pertama yang paling berkontribusi menaikkan kadar persen massa hidrogen, diikuti oleh Cr, Sn, Zr, V, Co, dan Ti secara berurutan, serta Ni pada posisi terakhir. Penambahan atom mangan pada logam paduan hidrida akan menyebabkan kenaikan terhadap volume unit sel yang ada pada logam hidrida. Hal ini dipengaruhi oleh dua hal yaitu ukuran radius atom dari unsur mangan dan teori bahwa atom mangan bukan merupakan atom pembentuk hidrida karena afinitas antara mangan dengan hydrogen yang kecil. Hal tersebut akan menyebabkan kenaikan terhadap volume unit sel yang akan menyebabkan potensi kimia yang lebih kecil terhadap atom hydrogen yang berada pada paduan. Hal inilah yang pada akhirnya dapat menurunkan kecenderungan atom hydrogen untuk tersebas dari logam paduan,

sehingga menyebabkan kapasitas hydrogen yang ada menjadi lebih tinggi. Hal ini mendukung hasil prediksi bahwa atom Mangan menjadi kontributor utama dalam menaikkan kapasitas hydrogen logam hidrida. Di sisi lain, hasil prediksi menunjukkan bahwa atom Ni bersifat menurunkan persen massa hydrogen. Atom Ni memiliki volume unit sel yang kecil, hal ini akan menyebabkan terjadinya tekanan equilibrium yang tinggi di dalam sel, sehingga kecenderungan lepasnya atom hydrogen ke bentuk gas yang lebih tinggi. Hal ini juga ditambah dengan fakta bahwa sebenarnya atom Ni ditambahkan ke dalam paduan sebagai katalis yang berfungsi untuk menurunkan energy pembentukan dari logam hidrida, namun dengan konsekuensi bahwa jumlah kapasitas hydrogen yang dimiliki hidrida tersebut juga menurun. Oleh karena itu dapat dikatakan bahwa prediksi atom nikel sebagai unsur yang bersifat menurunkan persen massa hydrogen sudah sesuai dengan teori yang ada.

b. Klasifikasi Berdasarkan Feature Importance Decision Tree

Tabel 4.9 Terjemahan Decision Tree untuk Persen Massa Hidrogen

Rank	Unsur	Batas 1	Efek 1	Batas 2	Efek 2
1	Mn	$>0,158$	Menaikkan	$\leq 0,158$	Menurunkan
2	Sn	$\leq 0,009$	Menurunkan	$>0,009$	Menaikkan
3	Co	$\leq 0,15$	Menaikkan	$>0,15$	Menurunkan
4	Ni	$\leq 0,12$	Menaikkan	$>0,12$	Menurunkan
4	Zr	$\leq 0,641$	Menurunkan	$>0,641$	Menaikkan
5	Ti	$\leq 0,380$	Menurunkan	$>0,380$	Menaikkan
5	V	$\leq 0,300$	Menaikkan	$>0,300$	Menurunkan

Klasifikasi berdasarkan decision tree memberikan hasil yang sama dengan klasifikasi regresi linear dengan atom Mn sebagai unsur yang berkontribusi menaikkan persen massa hidrogen dengan pengaruh tertinggi. Namun terdapat perbedaan untuk unsur yang berkontribusi menurunkan persen massa hidrogen, dimana unsur Co menempati urutan pertama, sedangkan unsur Ni menempati

urutan kedua. Cobalt pada praktiknya juga banyak digunakan sebagai katalis dalam proses hidrogenasi, hal inilah yang dapat menyebabkan menurunnya kapasitas hidrogen dari logam hidrida dengan paduan cobalt, sama halnya dengan nikel.

4.4. Modelling Variabel dengan Unsur-Unsur yang Dikelompokkan menjadi A dan B

Didapatkan hasil evaluasi dari masing-masing model yaitu regresi linear multivariate, decision tree, dan random forest untuk variabel dengan unsur-unsur yang dikelompokkan menjadi A dan B. Pengaruh rasio B/A terhadap setiap variabel output didapatkan untuk mengetahui pengaruhnya terhadap sifat-sifat penyimpanan hydrogen logam hidrida.

4.4.1. Hasil Multivariate Regression

Tabel 4.10 Hasil Regresi Multivariate untuk unsur A dan B yang dikelompokkan

Metric	Absorpsi	Absorpsi	Phase Abundance	Persen Massa H
R ² Train	0,046	0,03	0,067	0,166
R ² Validasi	-0,048	-0,109	-0,001	0,163
R ² Test	0,119	0,112	0,135	0,167
R ² Rata-rata	0,049	0,011	0,067	0,165
Mean Absolute Error	6,8	9,91	31,9	0,194
Root Mean Square Error	8,64	15,4	36,7	0,28

Regresi multivariate untuk unsur A dan B yang dikelompokkan menghasilkan kemampuan prediksi yang sangat rendah untuk keempat variabel output yaitu 0,049, 0,011, 0,067 dan 0,165 untuk selisih entalpi absorpsi, selisih entropi absorpsi,

phase abundance, and persen massa hydrogen secara berurutan. Nilai error yang dihasilkan pun cukup tinggi bagi selisih entalphi absorbs, selisih entropi absorbs, dan phase abundance. Hal ini menunjukkan bahwa dataset dengan unsur A dan B yang dikelompokkan, tidak cocok untuk diprediksi menggunakan multivariate regression dan model telah gagal memprediksi variable output yang diinginkan.

4.4.2. Hasil Decision Tree

Tabel 4.11 Hasil Regresi Decision Tree untuk unsur A dan B yang dikelompokkan

Metric	Absorpsi	Absorpsi	Phase Abundance	Persen Massa H
R ² Train	0,873	0,815	0,853	0,735
R ² Validasi	0,662	0,273	0,192	0,145
R ² Test	0,626	0,370	0,637	-0,825
R ² Rata-rata	0,720	0,486	0,560	0,018
Mean Absolute Error	4,36	30,8	22,4	0,267
Root Mean Square Error	5,75	33,6	25,6	0,414

Regresi yang dilakukan dengan model decision tree menghasilkan kemampuan prediksi pada 0,720, 0,486, 0,560, dan 0,018 untuk delta entalphi, delta entropi, phase abundance, dan persen massa hidrogen secara berurutan. Prediksi untuk delta entalphi tergolong tinggi dengan skor R² rata-rata diatas 0,70, sedangkan prediksi ketiga variabel output yang lain masih tergolong rendah karena berada dibawah angka 0,70. Penurunan skor R² validasi yang cukup signifikan terjadi pada prediksi , delta entropi, phase abundance, dan persen massa hidrogen yang menandakan overfitting yang cukup besar pada prediksi output tersebut. Mean Absolute error yang didapatkan berada pada nilai

4,36, 30,8, 22,4, dan 0,267 untuk delta entalpi, delta entropi, phase abundance, dan persen massa hidrogen secara berurutan. Sedangkan root mean square error berada pada angka 5,75, 33,6, 25,6, dan 0,414 untuk delta entalpi, delta entropi, phase abundance, dan persen massa hidrogen secara berurutan. Delta entalpi dan persen massa hidrogen menghasilkan nilai error yang cukup kecil karena mendekati nilai nol, hal yang sebaliknya terjadi pada error delta entropi dan phase abundance. Nilai error yang cukup besar pada delta entropi dan phase abundance ini dapat terjadi karena pada proses data mining terdapat banyak nilai delta entropi yang diasumsikan sebagai 110 KJ/mol untuk menemukan nilai delta entalpi. Hal inilah yang mungkin menyebabkan model tidak dapat secara pasti memprediksi nilai 110 KJ/mol sehingga nilai error yang terjadi menjadi cukup tinggi. Sedangkan phase abundance memiliki nilai yang cenderung sangat tinggi dan sangat rendah karena ada logam paduan yang memiliki nilai phase abundance 100% dan 0% sehingga cukup sulit bagi model untuk meraih nilai yang sangat tinggi maupun rendah tersebut.

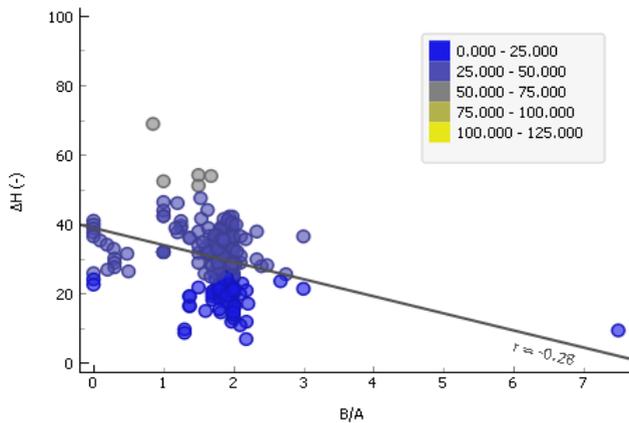
4.4.3. Hasil Random Forest

Tabel 4.12 Hasil Regresi Random Forest untuk unsur A dan B yang dikelompokkan

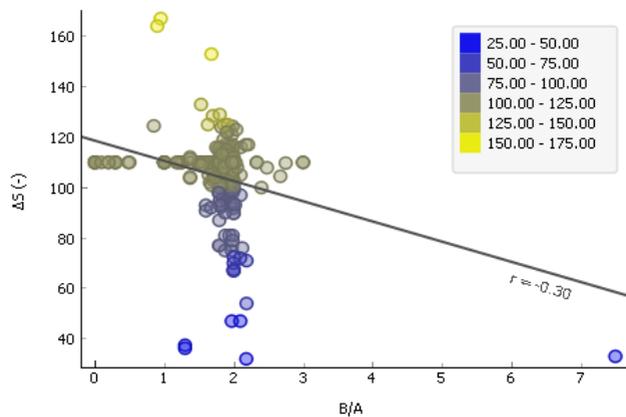
Metric	Absorpsi	Absorpsi	Phase Abundance	Persen Massa H
R ² Train	0,792	0,722	0,807	0,686
R ² Validasi	0,732	0,423	0,317	0,069
R ² Test	0,676	-0,46	-0,711	-0,843
R ² Rata-rata	0,73	0,238	0,137	-0,029
Mean Absolute Error	28,7	11,2	28,1	0,0325
Root Mean Square Error	29,6	19,8	48,3	0,067

Regresi random forest menghasilkan kemampuan yang lebih rendah apabila dibandingkan dengan regresi decision tree pada bagian sebelumnya untuk variable delta entropi, phase abundance, dan persen massa hidrogen pada nilai R^2 rata-rata 0,238, 0,136, dan -0,029 secara berurutan. Sedangkan prediksi delta entalphi menghasilkan kemampuan yang lebih tinggi dibandingkan dengan model yang lain pada bagian sebelumnya di angka 0,73. Ovefitting yang cukup besar terjadi pada prediksi delta entropi, phase abundance, dan persen massa hidrogen yang ditandai dengan menurunnya nilai R^2 pada data validasi dibandingkan dengan data training yang cukup signifikan. Mean Absolute error yang didapatkan berada pada nilai 28,7, 11,2, 28,1, dan 0,0325 untuk delta entalphi, delta entropi, phase abundance, dan persen massa hidrogen secara berurutan. Sedangkan root mean square error berada pada angka 29,6, 19,8, 48,3, dan 0,067 untuk delta entalphi, delta entropi, phase abundance, dan persen massa hidrogen secara berurutan. Dapat disimpulkan bahwa model random forest cocok untuk memprediksi nilai delta entalphi pada dataset unsur yang dikelompokkan menjadi A dan B.

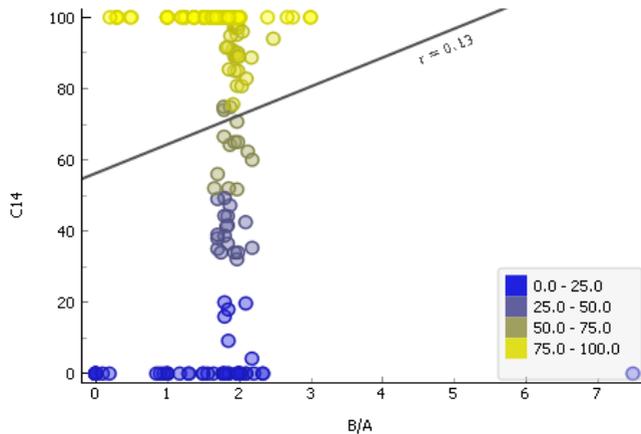
4.4.4. Pengaruh Rasio B/A terhadap Variabel Output



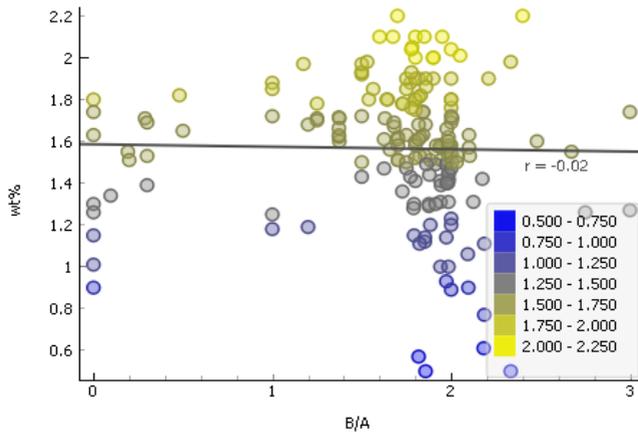
Gambar 4.19 Pengaruh rasio B/A terhadap variable delta entalphi



Gambar 4.20 Pengaruh rasio B/A terhadap variable delta entropi



Gambar 4.21 Pengaruh rasio B/A terhadap variable phase abundance C14



Gambar 4.22 Pengaruh rasio B/A terhadap variable persen massa hidrogen

Pengaruh rasio B/A ditemukan menggunakan regresi linear. Nilai B/A berbanding terbalik dengan nilai delta entalphi dan delta entropi dengan nilai koefisien $-0,38$ dan $-0,30$, sedangkan nilai B/A berbanding lurus dengan nilai phase abundance C14

dengan nilai koefisien 0,13. Nilai persen massa hydrogen terlihat tidak terlalu terpengaruh oleh perubahan nilai B/A dimana grafik regresi menunjukkan garis yang cukup mendatar dengan koefisien yang cukup kecil pada -0,02.

Hubungan antara rasio B/A dan nilai entalpi serta entropi menunjukkan bahwa energi pembentukan akan menurun ketika nilai B menjadi lebih besar dan nilai A menjadi lebih kecil. Atom A pada dasarnya akan membentuk hidrida yang sangat stabil, namun dengan konsekuensi bahwa temperature serta tekanan yang dibutuhkan untuk pembentukan hidrida tersebut akan menjadi sangat tinggi yang berarti bahwa energy pembentukan yang dibutuhkan akan menjadi sangat tinggi pula. Hal inilah sebenarnya yang menjadi dasar penambahan atom B kepada logam paduan AB_2 dimana secara teoritis logam B akan membentuk hidrida yang tidak stabil namun dengan temperature dan tekanan pembentukan yang rendah sehingga hal ini akan menurunkan energy pembentukan dari logam paduan hidrida AB_2 . Hal ini secara sukses ditunjukkan oleh modelling hubungan antara rasio B/A dengan energi pembentukan yang telah didapatkan. Hubungan antara rasio B/A dan phase abundance menunjukkan kenaikan fasa C14 seiring dengan nilai atom B yang bertambah besar atau atom A yang bertambah kecil. Penambahan atom B akan menyebabkan penambahan pada nilai e/a yang berkontribusi untuk menentukan terbentuknya fasa C14 atau C15, dimana penambahan nilai e/a akan mendorong terbentuknya fasa C15 dan menghambat terbentuknya fasa C14 sehingga modelling yang dihasilkan tidak sesuai dengan teori yang ada. Namun, hal ini dapat dijelaskan apabila grafik plot untuk phase abundance dilihat, dimana terdapat banyak nilai yang berada pada angka 0% dan 100% sedangkan nilai yang lain membentuk garis yang cukup horizontal sehingga hal inilah yang bisa dikatakan mempengaruhi hasil modelling yang didapatkan.

Hubungan antara rasio B/A dan persen massa hydrogen menunjukkan bahwa rasio B/A tidak berpengaruh terhadap perubahan persen massa hydrogen. Perubahan nilai persen massa hydrogen dipengaruhi oleh ukuran radius unit sel, dimana ukuran radius unit sel lebih bergantung kepada jenis unsur paduan secara individual dibandingkan dengan jumlah unsur paduan yang dikelompokkan menjadi unsur A maupun B.

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

1. Feature Engineering, pengecekan null, uji multikolinearitas, deteksi dan penanganan outliers, data scaling, dan pemilihan jumlah tree optimum untuk model random forest dilakukan sebagai processing data awal.
2. Modelling variable dengan unsur-unsur yang tidak dikelompokkan memberikan hasil terbaik pada model random forest untuk seluruh variable output dengan nilai R^2 yang tertinggi serta overfitting yang paling rendah.
3. Klasifikasi pengaruh masing masing unsur terhadap variable output memberikan Ni sebagai unsur terbaik untuk menurunkan delta entalphi dan delta entropi, kromium sebagai unsur terbaik untuk menaikkan kadar fasa C14, dan Mn sebagai unsur terbaik untuk menaikkan persen massa hydrogen.
4. Modelling variable dengan unsur-unsur yang dikelompokkan menjadi unsur A dan B memberikan hasil terbaik pada model decision tree untuk variable delta entropi, phase abundance, dan persen massa hydrogen, serta model random forest untuk variable delta entalphi
5. Klasifikasi pengaruh rasio B/A terhadap variable output memberikan hasil yang berbanding terbalik untuk delta entalphi dan delta entropi, hasil yang berbanding lurus untuk phase abundance C14, dan tidak berpengaruh terhadap persen massa hydrogen

5.2. Saran

1. Proses data mining untuk membentuk dataset sebaiknya dilakukan dengan memperhatikan prosedur eksperimen

yang ada pada setiap jurnal agar tidak terjadi outlier yang terlalu parah.

2. Disarankan untuk membangun sebuah template yang sama bagi seluruh model pada saat dilakukan fitting di software Python.
3. Penyeleksian fitur harus dilakukan dengan hati-hati agar hubungan antara seluruh fitur dapat dikoneksikan dengan baik oleh model machine learning.
4. Disarankan untuk melakukan verifikasi hasil prediksi dengan logam paduan yang berasal dari eksperimen yang baru guna membuktikan kebenaran hasil prediksi model machine learning.

DAFTAR PUSTAKA

- [1] U. Ulmer *et al.*, “Cost reduction possibilities of vanadium-based solid solutions – Microstructural, thermodynamic, cyclic and environmental effects of ferrovandium substitution,” *J. Alloys Compd.*, vol. 648, pp. 1024–1030, Nov. 2015, doi: 10.1016/j.jallcom.2015.07.110.
- [2] S. Suwarno and A. Shahab, “Prediksi Derajat Presipitasi Karbida Krom pada Baja Tahan Karat Austenitik dengan Metode Jaringan Syaraf Tiruan,” 2004, [Online]. Available: https://www.researchgate.net/publication/314256935_Prediksi_Derajat_Presipitasi_Karbida_Krom_pada_Baja_Tahan_Karat_Austenitik_dengan_Metode_Jaringan_Syaraf_Tiruan.
- [3] A. Rahnama, G. Zepon, and S. Sridhar, “Machine learning based prediction of metal hydrides for hydrogen storage, part I: Prediction of hydrogen weight percent,” *Int. J. Hydrog. Energy*, vol. 44, no. 14, pp. 7337–7344, Mar. 2019, doi: 10.1016/j.ijhydene.2019.01.261.
- [4] A. Züttel, “Materials for hydrogen storage,” *Mater. Today*, vol. 6, no. 9, pp. 24–33, Sep. 2003, doi: 10.1016/S1369-7021(03)00922-2.
- [5] L. Schlapbach and A. Züttel, “Hydrogen-storage materials for mobile applications,” *Nature*, vol. 414, no. 6861, pp. 353–358, Nov. 2001, doi: 10.1038/35104634.
- [6] C. M. Graça Araújo, “Hydrogen storage materials: design, catalysis, thermodynamics, structure and optics,” *Acta Universitatis Upsaliensis*, Uppsala, 2008.
- [7] M. Hirscher, Ed., *Handbook of hydrogen storage: new materials for future energy storage*. Weinheim: Wiley-VCH, 2010.
- [8] A. A. Volodin *et al.*, “Study of hydrogen storage and electrochemical properties of AB₂-type Ti_{0.15}Zr_{0.85}La_{0.03}Ni_{1.2}Mn_{0.7}V_{0.12}Fe_{0.12} alloy,” *J.*

- Alloys Compd.*, vol. 793, pp. 564–575, Jul. 2019, doi: 10.1016/j.jallcom.2019.03.134.
- [9] G. Sandrock, “A panoramic overview of hydrogen storage alloys from a gas reaction point of view,” *J. Alloys Compd.*, vol. 293–295, pp. 877–888, Dec. 1999, doi: 10.1016/S0925-8388(99)00384-9.
- [10] F. Feng, “Electrochemical behaviour of intermetallic-based metal hydrides used in Ni/metal hydride (MH) batteries: a review,” *Int. J. Hydrog. Energy*, vol. 26, no. 7, pp. 725–734, Jul. 2001, doi: 10.1016/S0360-3199(00)00127-0.
- [11] L. Klebanoff, *Hydrogen Storage Technology: Materials and Applications*. 2016.
- [12] D. P. Broom, *Hydrogen storage materials: the characterisation of their storage properties*. London ; New York: Springer, 2011.
- [13] E. Alpaydin, *Introduction to machine learning*, 2nd ed. Cambridge, Mass: MIT Press, 2010.
- [14] G. S. Linoff and M. J. A. Berry, *Data mining techniques: for marketing, sales, and customer relationship management*, 3rd ed. Indianapolis, IN: Wiley Pub, 2011.
- [15] H. Jiawie, *Data Mining*. Elsevier, 2012.
- [16] J. R. Hattrick-Simpers, K. Choudhary, and C. Corngale, “A simple constrained machine learning model for predicting high-pressure-hydrogen-compressor materials,” *Mol. Syst. Des. Eng.*, vol. 3, no. 3, pp. 509–517, 2018, doi: 10.1039/C8ME00005K.
- [17] M. Witman *et al.*, “Extracting an Empirical Intermetallic Hydride Design Principle from Limited Data via Interpretable Machine Learning,” *J. Phys. Chem. Lett.*, vol. 11, no. 1, pp. 40–47, Jan. 2020, doi: 10.1021/acs.jpcclett.9b02971.
- [18] S.-R. Kim, K.-Y. Lee, and J.-Y. Lee, “Improved low-temperature dischargeability of C14-type Zr-Cr-Ni Laves phase alloy,” *J. Alloys Compd.*, vol. 223, no. 1, pp. 22–27, May 1995, doi: 10.1016/0925-8388(94)01499-X.

- [19] K.-H. Young, J. Nei, C. Wan, R. Denys, and V. Yartys, “Comparison of C14- and C15-Predominated AB₂ Metal Hydride Alloys for Electrochemical Applications,” *Batteries*, vol. 3, no. 4, p. 22, Jul. 2017, doi: 10.3390/batteries3030022.

Halaman ini sengaja dikosongkan

LAMPIRAN

Lampiran A. Hasil Regresi Setiap Algoritma

Tabel A.1 Hasil Regresi Linear Multivariate untuk Setiap Variabel Output

Metric	Absorpsi	Absorpsi	Phase Abundance	Persen Massa H
Kemampuan Prediksi - R ²	0.367	0.177	0.448	0.667
Mean Absolute Error	4,26	10,2	10,1	0,111
Root Mean Square Error	6,15	17,1	13,4	0,148

Tabel A.2 Hasil Regresi Decision Tree untuk Setiap Variabel Output

Metric	Absorpsi	Absorpsi	Phase Abundance	Persen Massa H
Kemampuan Prediksi - R ²	0.346	-0.16	0.785	0.498
Mean Absolute Error	5,97	7,57	9,18	0,0967
Root Mean Square Error	7,77	14,2	11,3	0,161

Tabel A.3 Hasil Regresi Random Forest untuk Setiap Variabel Output

Metric	Absorpsi	Absorpsi	Phase Abundance	Persen Massa H
---------------	-----------------	-----------------	------------------------	-----------------------

Kemampuan Prediksi - R ²	0,647	0,554	0,832	0,668
Mean Absolute Error	4,36	30,8	22,4	0,101
Root Mean Square Error	5,75	33,6	25,6	0,126

Lamiran B. Pengaruh Komposisi Kimia Terhadap Sifat Penyimpanan Hidrogen

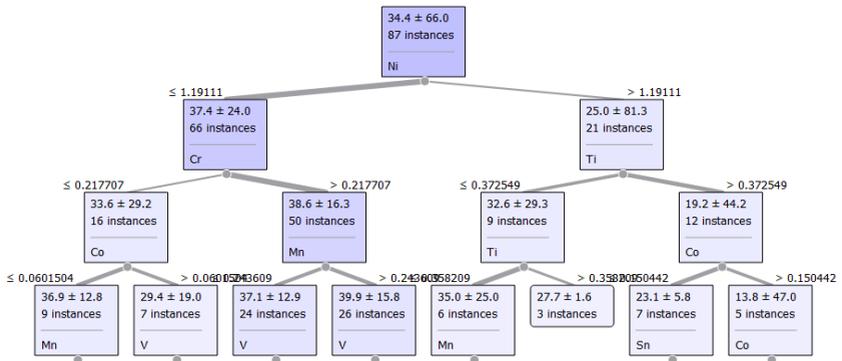
Tabel B.1 Hasil Regresi Pengaruh Komposisi Kimia Terhadap Selisih Entalpi Absorpsi

Unsur	Nilai r	Pengaruh terhadap entalpi absorpsi
Ti	-0.21	Menurunkan
Zr	0.21	Menaikkan
Mn	0.42	Menaikkan
Co	-0.05	Menurunkan
Cr	0.43	Menaikkan
V	-0.07	Menurunkan
Ni	-0.59	Menurunkan
Sn	0.22	Menaikkan

Tabel B.2 Ranking Pengaruh Komposisi Kimia terhadap Selisih Entalpi Absorpsi berdasarkan Regresi Linear

Rangking	Unsur	Skor	Fungsi
1	Ni	-0.59	Menurunkan
2	Ti	-0.21	Menurunkan
3	V	-0.07	Menurunkan
4	Co	-0.05	Menurunkan
5	Zr	0.21	Menaikkan
6	Sn	0.22	Menaikkan
7	Mn	0.42	Menaikkan

8	Cr	0.43	Menaikkan
---	----	------	-----------



Gambar B.1 Klasifikasi Pengaruh Komposisi Kimia terhadap Selisih Entalpi Absorpsi berdasarkan Decision Tree

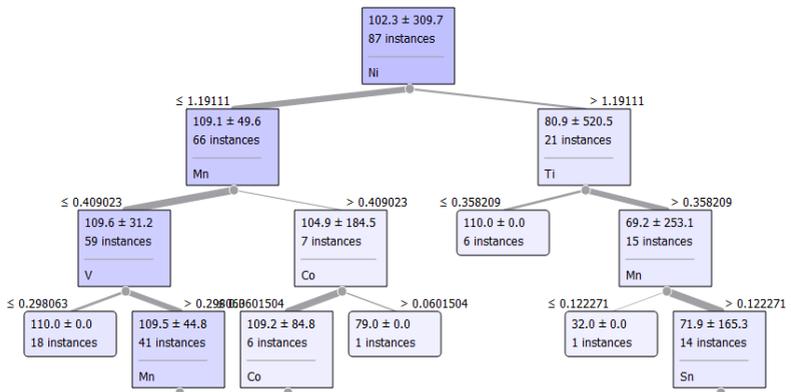
Tabel B.3 Hasil Regresi Pengaruh Komposisi Kimia Terhadap Selisih Entropi Absorpsi

Komposisi	Nilai R	Pengaruh terhadap entropi absorpsi
Ti	-0.27	Menurunkan
Zr	0.27	Menaikkan
Mn	0.37	Menaikkan
Co	0.04	Menaikkan
Cr	0.27	Menaikkan
V	-0.11	Menurunkan
Ni	-0.54	Menurunkan
Sn	0.06	Menaikkan

Tabel B.4 Ranking Pengaruh Komposisi Kimia terhadap Selisih Entropi Absorpsi berdasarkan Regresi Linear

Rangking	Komposisi	Skor	Fungsi
----------	-----------	------	--------

1	Ni	-0.54	Menurunkan
2	Ti	-0.27	Menurunkan
3	V	-0.11	Menurunkan
4	Co	0.04	Menaikkan
5	Sn	0.06	Menaikkan
6	Zr	0.27	Menaikkan
7	Cr	0.27	Menaikkan
8	Mn	0.37	Menaikkan



Gambar B.2 Klasifikasi Pengaruh Komposisi Kimia terhadap Selisih Entropi Absorpsi berdasarkan Decision Tree

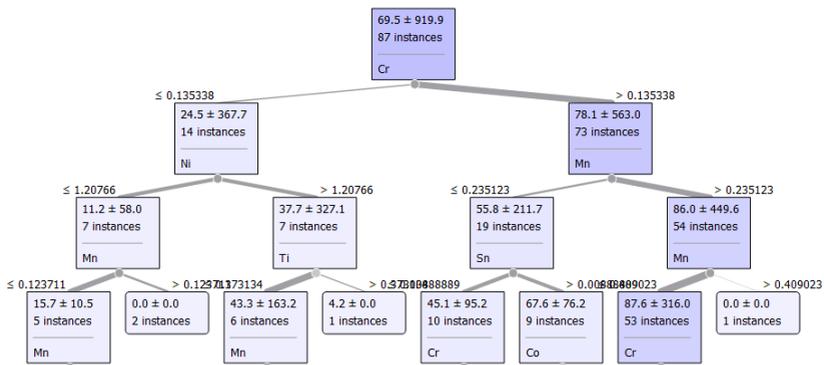
Tabel B.5 Hasil Regresi Pengaruh Komposisi Kimia Terhadap Phase Abundance C14

Komposisi	Nilai R	Pengaruh terhadap fasa C14
Ti	-0.37	Menurunkan
Zr	0.37	Menaikkan
Mn	0.73	Menaikkan
Co	-0.18	Menurunkan
Cr	0.46	Menaikkan
V	0.35	Menaikkan

Ni	-0.55	Menurunkan
Sn	0.54	Menaikkan

Tabel B.6 Ranking Pengaruh Komposisi Kimia terhadap Phase Abundance C14 berdasarkan Regresi Linear

Rangking	Komposisi	Skor	Fungsi
1	Mn	0.73	Menaikkan
2	Sn	0.54	Menaikkan
3	Cr	0.46	Menaikkan
4	Zr	0.37	Menaikkan
5	V	0.35	Menaikkan
6	Co	-0.18	Menurunkan
7	Ti	-0.37	Menurunkan
8	Ni	-0.55	Menurunkan



Gambar B.3 Klasifikasi Pengaruh Komposisi Kimia terhadap Phase Abundance C14 berdasarkan Decision Tree

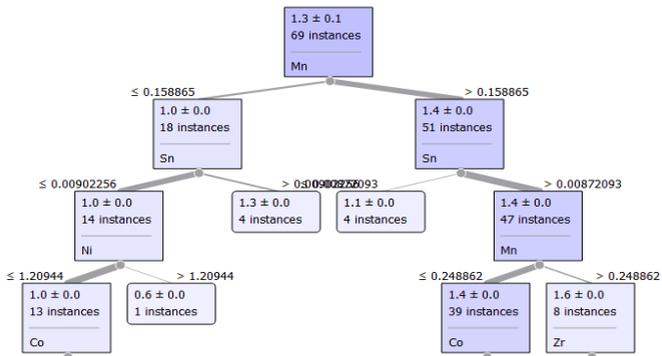
Tabel B.7 Hasil Regresi Pengaruh Komposisi Kimia Terhadap Persen Massa Hidrogen

Komposisi	Nilai R	Pengaruh terhadap wt%
-----------	---------	-----------------------

Ti	-0.22	Menurunkan
Zr	0.22	Menaikkan
Mn	0.63	Menaikkan
Co	-0.21	Menurunkan
Cr	0.53	Menaikkan
V	0.04	Menaikkan
Ni	-0.47	Menurunkan
Sn	0.44	Menaikkan

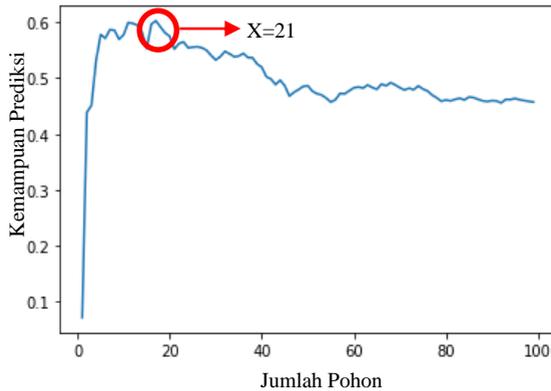
Tabel B.8 Ranking Pengaruh Komposisi Kimia terhadap Persen Massa Hidrogen berdasarkan Regresi Linear

Rangking	Komposisi	Skor	Fungsi
1	Mn	0.63	Menaikkan
2	Cr	0.53	Menaikkan
3	Sn	0.44	Menaikkan
4	Zr	0.22	Menaikkan
5	V	0.04	Menaikkan
6	Co	-0.21	Menurunkan
7	Ti	-0.22	Menurunkan
8	Ni	-0.47	Menurunkan

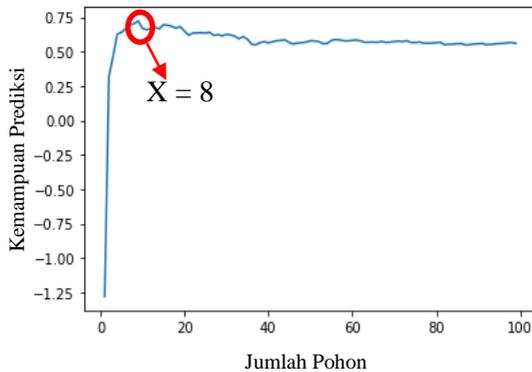


Gambar B.4 Klasifikasi Pengaruh Komposisi Kimia terhadap Persen Massa Hidrogen berdasarkan Decision Tree

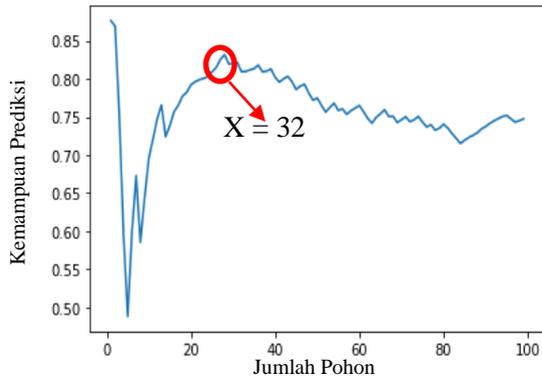
Lampiran C. Pemilihan jumlah tree untuk model Random Forest



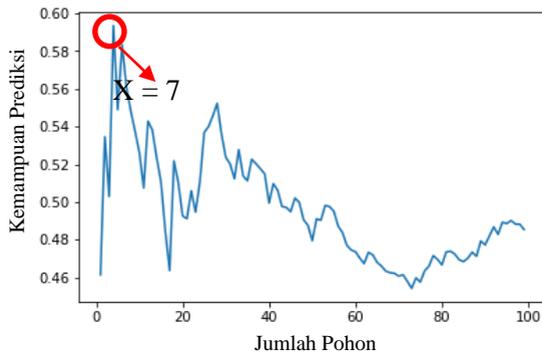
Gambar C.1 Pengecekan Jumlah Tree Optimum untuk Variabel Selisih Entalphi Pembentukan



Gambar C.2 Pengecekan Jumlah Tree Optimum untuk Variabel Selisih Entropi Pembentukan



Gambar C.3 Pengecekan Jumlah Tree Optimum untuk Variabel Phase Abundance C14



Gambar C.4 Pengecekan Jumlah Tree Optimum untuk Variabel Persen Massa Hidrogen

BIODATA PENULIS



Penulis dilahirkan di Denpasar, 29 Maret 1998, merupakan anak pertama dari 2 bersaudara dari pasangan Yackie dan Dewi Safitri. Penulis telah menempuh pendidikan formal di TK Pembina Negeri Denpasar, MIT Tawakal, SMP N 2 Tabanan, dan SMA N 1 Tabanan. Setelah lulus dari SMA N 1 Tabanan pada tahun 2016, Penulis mengikuti SNMPTN dan diterima di Departemen Teknik Mesin FTI-ITS pada tahun yang sama dan terdaftar dengan NRP 0211164000013.

Di Departemen Teknik Mesin ini Penulis mengambil Bidang Studi Metalurgi. Penulis sempat aktif menjabat sebagai Presiden di ITS Debating Society, Volunteer di ITS International Office, serta berhasil meraih berbagai penghargaan tingkat nasional dan internasional. Penulis juga berkesempatan untuk mengikuti program pertukaran pelajar ke Turki serta Taiwan.