

## BAB I PENDAHULUAN

### 1.1. Latar Belakang

Era Revolusi Industri 4.0 juga ditandai dengan meningkatnya tren otomatisasi, *Internet of Things (IoT)*, *Big Data*, dan Teknologi *Cloud*. Era Industri 4.0 menawarkan peluang untuk pekerjaan ilmu data. Sangat penting untuk dipahami bahwa semua komponen Industri 4.0, seperti *IoT*, *Big Data*, *AI*, *Machine Learning*, dan *Cloud Computing* membutuhkan data. Tanpa data, komponen tersebut tidak dapat menjawab kebutuhan masyarakat secara akurat (Arfienda, 2018). Salah satu aplikasi yang saat ini dapat digunakan untuk *machine learning* adalah *Software R Studio*. Aplikasi ini menyertakan bahasa pemrograman R sebagai logika keluaran program untuk memberikan wawasan atau rekomendasi yang tepat.

Dalam dunia *Data Mining* dan analisis data, *Clustering* merupakan salah satu teknik yang paling penting bagi PT. Integral Industrial Indonesia digunakan untuk membagi data menjadi beberapa kelompok berdasarkan kesamaan fitur atau atribut. Hal ini berguna untuk melakukan analisis jenis data UID, XYZ PEAK, XYZ RMS, RATE. yang akan di kelompokkan berdasarkan batas *Threshold*. PT. Integral Industrial Indonesia ini pada tahun 2022 memiliki banyak jenis data dengan tingkat data yang melebihi Outlier sangat banyak, dan melebihi batas *Threshold*. Tujuan pada penelitian ini meningkatkan data efektivitas kinerja *Sensor Vibration* yang ditenpelkan pada *gearbox* mesin kertas, dan mengambil keputusan yang lebih akurat. Oleh karena itu, penelitian tentang *Clustering* menjadi sangat penting untuk dilakukan, terutama di PT. Integral Industrial Indonesia yang masih kurang memanfaatkan teknik *Clustering* ini secara optimal.

Dalam penelitian ini *Clustering*, terdapat beberapa metode *Clustering* yang umum digunakan, seperti *K-Means*, *Hierarchical Clustering*. Setiap metode memiliki kelebihan dan kekurangan masing-masing. Oleh karena itu, penelitian *Clustering* perlu dilakukan untuk membandingkan keefektifan metode-metode tersebut dalam memecahkan masalah pada PT. Integral Industrial Indonesia. Selain itu, penggunaan teknik *Clustering* juga sangat bergantung pada pemilihan parameter yang tepat, seperti jumlah *Cluster* atau nilai Euclidean. Oleh karena itu,

penelitian *Clustering* juga perlu mempelajari teknik pemilihan parameter yang tepat agar teknik *Clustering* dapat dioptimalkan dengan baik pada perusahaan.

Dalam melakukan penelitian ini bekerja sama dengan PT. Integral Industrial Indonesia sangat penting dalam pengelompokkan data. PT. Integral Industrial Indonesia adalah perusahaan yang beroperasi di lingkungan industri yang kompleks di mana perencanaan, pengoperasian, dan pemeliharaan aset yang optimal merupakan hal yang sangat penting. Mereka mengembangkan *platform* perangkat lunak khusus untuk mengotomatiskan pekerjaan data dan membangun keunggulan kompetitif yang berkelanjutan dengan menghasilkan pengembalian yang tinggi selama masa pakai aset.

Proses pengumpulan data berlangsung selama 52 hari dari tanggal 12 Januari 2023 sampai dengan 4 Maret 2023. Penelitian ini dilakukan dengan menggunakan teknik analisis machine learning antara lain model *Clustering K-Means*, *Hierarichal Clustering*, untuk mendeteksi batas *Treshold*. Alat mesin berdasarkan data *Vibration* yang dihasilkan selama proses manufaktur. Tujuan dari karya penelitian ini adalah memberikan solusi bagi PT. Integral Industrial Indonesia untuk meningkatkan produksi mesin kertas dan mencegah kerusakan pada komponen mesin kertas yang tidak terduga.

## 1.2. Perumusan Masalah

1. Bagaimana penggunaan metode *Clustering K – Means* dan *Hierarichal Clustering* dalam mengelompokkan data *JSON file* di PT. Integral Industrial Indonesia?
2. Bagaimana teknik *Clustering* dapat digunakan untuk menganalisis data pada sensor *Vibration* yang di tempelkan pada *Gearbox* berdasarkan data sensor?
3. Bagaimana mengoptimalkan kinerja model *Clustering K-Means* dan *Hierarichal Clustering* dengan teknik *Machine Learning*?

### 1.3. Tujuan Penelitian

Tujuan penelitian yang dilakukan di PT. Integral Industrial Indonesia adalah sebagai berikut:

1. Mampu mengidentifikasi pola atau karakteristik data JSON *file* yang ada pada sensor yang terkumpul.
2. Menentukan metode apa yang sebaiknya diterapkan dalam pengelompokan data di PT. Integral Industrial Indonesia.
3. Mampu mengelompokkan data sensor dengan metode *Clustering K-Means*, *Hierarichal Clustering*.
4. Mampu mengetahui manfaat dan keuntungan yang diperoleh perusahaan dalam penelitian ini, dari segi aspek finansial.

### 1.4. Batasan Masalah

Pada penelitian yang dilakukan, perlu adanya pembatasan agar penulisan lebih terarah, tidak meluas dan tidak menyimpang dari tujuan penulisan. Pembatasan tersebut antara lain sebagai berikut :

1. Membatasi objek yang akan di-cluster, seperti data numerik, data teks, yang ada pada data JSON *file* di PT. Integral Industrial Indonesia.
2. Menggunakan data JSON yang akan di *Clustering K – Means* guna untuk mengelompokkan data dengan karakteristik yang sama ke satu wilayah yang sama dan data dengan karakterisitik yang berbeda ke wilayah lain.
3. Membatasi parameter guna mengukur *Treshold* data pada sensor di PT. Integral Industrial Indonesia.
4. Data yang diperoleh dalam pembuatan model sebanyak 1262 data yaitu data UID, *XYZ peak*, *XYZ rms*, dan *Rate*.

## 1.5.State Of The Art

Tabel 1.1 State Of The Art

Judul Jurnal	Pembahasan
<p><b>Judul</b> Pengelompokan Mahasiswa Potensial <i>Drop Out</i> menggunakan Metode <i>Clustering K-Means</i>.</p> <p><b>Peneliti</b> Ieannoal Vhallaha, Sumijanb, Julius Santonyc.</p> <p><b>Lokasi</b> Padang, Indonesia</p> <p><b>Tahun</b> 2018</p>	<p><u>Hasil Penelitian</u> Makalah penelitian menjelaskan bahwa tujuan dari penelitian ini adalah untuk mengelompokkan siswa yang berisiko putus sekolah menurut berbagai karakteristik termasuk sistem poin semester kumulatif, IPK kumulatif dan semester kumulatif. Proses pengelompokan K-Kean dilakukan dengan memberikan titik pusat awal secara acak kepada sekelompok siswa. Jadi, hasil yang diperoleh dengan metode K adalah untuk siswa yang memiliki kemungkinan putus sekolah. Kelas 2014 berada di cluster 0 dengan jumlah 4 siswa atau 30,77% dari 13 sampel, kelas 2015 berada di cluster 1 dengan jumlah 4 siswa dan sebanyak 2 siswa di cluster 2 atau 66,7% dari 9 dari 10 sampel, total 10 kelas. 10 siswa, atau 50% dari 24 sampel, dan 20 17 kelas berada di Cluster 2, yang memiliki kekuatan 4 siswa, atau 22,22% dari 18 sampel.</p> <p><u>Alasan menjadi tinjauan penelitian</u> Dengan menggunakan jurnal berikut dapat mengetahui bahwa Dengan menggunakan Metode <i>Clustering K-Means</i> untuk memprediksi Total SKS, IPK, dan dapat mendeteksi mahasiswa yang berpotensi Drop Out.</p>
<p><b>Judul</b> Penerapan Data Mining Menentukan Strategi Penjualan Variasi Mobil Menggunakan Metode <i>K-Means Clustering</i>.</p> <p><b>Peneliti</b> Haditsah Annur</p> <p><b>Lokasi</b></p>	<p>Hasil penelitian Dalam penelitian ini, analisis data mining dilakukan dengan menggunakan metode k dan teknik clustering. Penggunaan <i>K-Method</i> pada penelitian ini dikarenakan K-Method merupakan metode pengelompokan data non-hierarkis (blok), yang bertujuan untuk</p>

<p>Gorontalo, Indonesia</p> <p><b>Tahun</b> 2019</p>	<p>membagi data yang ada menjadi dua atau lebih kelompok. Hasil penelitian ini mengelompokkan data penjualan untuk mengetahui kemungkinan tanggal atau tren pembelian pelanggan. Dari hasil penelitian dapat diberikan saran dan pertimbangan dalam menentukan strategi pemasaran yaitu. sekaligus mengeliminasi produk riset dengan menentukan posisi cluster tertinggi.</p> <p><b><u>Alasan menjadi tinjauan penelitian</u></b></p> <p>Tujuan dari penelitian berikut digunakan menjadi tinjauan penelitian yaitu untuk mengetahui penggunaan model Clustering K-Means dengan metode yang memiliki akurasi persentase terbaik untuk mengetahui penjualan produk.</p>
<p><b>Judul</b> <i>Rain Prediction Clustering in Australia Using the K-Means Algorithm in the WEKA and RStudio Application.</i></p> <p><b>Peneliti</b> Dinar Ajeng Kristiyanti<sup>1</sup>, Irwansyah Saputra<sup>2</sup>, Rina<sup>3</sup></p> <p><b>Lokasi</b> Yogyakarta, Indonesia</p> <p><b>Tahun</b> 2021</p>	<p><b><u>Hasil Penelitian</u></b></p> <p>Tujuan dari penelitian jurnal ini adalah bagaimana cara menciptakan cluster yang ideal dalam memprediksi curah hujan di Australia berdasarkan presentase sum of squares error (SSE) menggunakan algoritme K-Means dengan aplikasi WEKA dan RStudio. Metode atau tahapan yang diterapkan dalam melakukan prediksi hujan di Australia yaitu melalui beberapa tahapan diantaranya Pengumpulan Data, Data <i>Pre-processing</i> (termasuk dilakukan penanganan Missing Value didalamnya), Pemodelan Data Mining dengan menerapkan algoritma <i>K-Means Clustering</i> menggunakan WEKA dan RStudio, Validasi hasil dengan SSE serta Visualiasi Data menggunakan plot. Berdasarkan hasil yang diperoleh, cluster yang berjumlah 2 dengan SSE 28.0% merupakan cluster ideal untuk memprediksi hujan di Australia. Pada software WEKA cluster hujan diwakili oleh node berwarna biru dan cluster tidak hujan diwakili oleh node</p>

	<p>berwarna merah. Sedangkan pada software RStudio cluster hujan diwakili oleh node berwarna hitam dan cluster tidak hujan diwakili oleh node berwarna merah.</p> <p><b><u>Alasan menjadi tinjauan penelitian</u></b>  Dengan menggunakan jurnal berikut dapat digunakan menjadi tinjauan penelitian karena model yang digunakan Mendapatkan cluster yang ideal dalam memprediksi curah hujan di Australia dengan membandingkan hasil yang didapatkan menggunakan aplikasi WEKA dan RStudio.</p>
<p><b>Judul</b>  Pengelompokan Titik Api di Provinsi Jambi dengan Algoritma <i>Agglomerative Hierarchical Clustering</i>.</p> <p><b>Peneliti</b>  Krisman Pratama Simanjuntak, Ulfa Khaira.</p> <p><b>Lokasi</b>  Indoneisa</p> <p><b>Tahun</b>  2021</p>	<p>Berdasarkan pengujian dan analisis algoritma <i>Agglomerative Hierarchical Clustering</i> untuk <i>Clustered Hotspots</i> di Provinsi Jambi tahun 2019 diuji dengan menggunakan metode <i>silhouette scoring</i></p> <p>Berdasarkan koefisien tersebut dapat disimpulkan bahwa algoritma <i>agglomerative hierarchical clustering</i> berhasil.</p> <p>diperkenalkan pada tahun 2019 dalam klasifikasi hot spot di Provinsi Jambi. Hasil klaster dievaluasi menggunakan metode koefisien siluet dan diperoleh 2 klaster optimal, dengan klaster pertama memiliki hingga 6283 titik, rata-rata kepercayaan 73,49642%, rata-rata kecerahan 325,6746, FRP rata-rata 679,95973, yang kedua memiliki 3,47 titik 9 dan yang kedua memiliki 3 titik. %, kecerahan rata-rata 405.0715, FRP rata-rata 679.95973. Oleh karena itu, probabilitas kebakaran di cluster 2 lebih tinggi daripada di cluster 1 dan oleh karena itu memerlukan perhatian dan pengelolaan yang lebih. Berdasarkan hasil clustering, daerah yang memiliki titik panas terbanyak adalah daerah Muaro Jambi. Ini tentang peristiwa nyata yang terjadi di tahun 2019.</p>

	<p><b><u>Alasan menjadi tinjauan penelitian</u></b> Alasan jurnal berikut menjadi tinjauan penelitian adalah sama-sama menggunakan model yang akan digunakan pada penelitian yang akan dilakukan yaitu <i>Agglomerative Hierarchical Clustering</i> dengan menggunakan perhitungan <i>Silhouette</i>.</p>
<p><b>Judul</b> Perbandingan Algoritma DBSCAN K-Means Untuk Pengelempokkan Madrasah Aliyah Provinsi Jawa Timur.</p> <p><b>Peneliti</b> EMIR AKBAR</p> <p><b>Lokasi</b> Indonesia</p> <p><b>Tahun</b> 2023</p>	<p><b><u>Hasil Penelitian</u></b> Pada penelitian jurnal berikut melakukan prediksi DBSCAN, penulis melakukan pengelompokan menggunakan metode K-Means Clustering yang dikembangkan dengan menggunakan metode DBSCAN Clustering yang akan dikembangkan dengan DBSCAN Clustering, Penelitian ini dilakukan karena metode K-Means Clustering memiliki beberapa kekurangan, yaitu tidak dapat mendeteksi data noise/outliers. Berdasarkan hasil Clustering yang dilakukan oleh penulis, pengembangan metode K-Means Clustering mendapatkan nilai <i>Silhouette Score</i> terbaik pada pembentukan 2 klaster terbaik yaitu 0,825 atau 82.5%. sedangkan hasil pembentukan kluster menggunakan K-Means Clustering mendapatkan nilai <i>Silhouette Score</i> terbaik pada pembentukan 2 klaster, yaitu 0.845 atau 84.5%. Kedua hasil tersebut ke dalam struktur yang kuat karena memiliki nilai di antara 0.7 dan 1.</p> <p><b><u>Alasan menjadi tinjauan penelitian</u></b> Dengan menggunakan jurnal berikut dapat digunakan menjadi tinjauan penelitian dengan menggunakan model DBSCAN Clustering.</p>

## 1.6.Sistematika Penulisan

Adapun sistematika penulisan laporan tugas akhir ini adalah sebagai berikut :

**BAB I : PENDAHULUAN**

Dalam bab ini dibahas latar belakang, rumusan masalah, tujuan penelitian, definisi masalah dan sistematika penulisan.

**BAB II : TINJAUAN PUSTAKA**

Landasan teori adalah teori-teori yang berkaitan dengan penelitian yang dilakukan oleh penulis, yang berfungsi sebagai analisis untuk menjelaskan fakta-fakta yang ada. Landasan teoritis untuk ini diperoleh dari hasil penelitian literatur pada subjek dan dijadikan sebagai dasar untuk pembahasan dan pemecahan masalah dalam tesis diploma ini.

**BAB III : METODOLOGI PENELITIAN**

Bab ini merupakan gambaran Langkah-langkah secara sistematis yang dilakukan penulisan dari awal hingga akhir penelitian sehingga pelaksanaan penelitian menjadi jelas dan terfokus sesuai dengan tujuan penelitian serta memberikan kemudahan bagi pembaca untuk memahami penelitian yang dilakukan.

**BAB IV : PENGUMPULAN DAN PENGOLAHAN DATA**

Berisikan data umum perusahaan dan data yang dibutuhkan untuk diolah sebagai dasar untuk memecahkan masalah yang dihadapi serta digunakan sebagai pengolahan data di dalam melakukan analisa, menarik kesimpulan dan pengajuan saran.

**BAB V : KESIMPULAN & SARAN**

Bab ini berisi kesimpulan dari temuan penelitian dan proposal untuk perusahaan. Saran yang disampaikan didasarkan pada hasil analisis penelitian untuk perbaikan di masa yang akan datang.