

# *Systematic Literature Review: Pengembangan Sistem Intelijen untuk Prediksi Suhu*

Dinda Jaelani Hidayat <sup>#1</sup>, G.S.Budhi Dharmawan <sup>#2</sup>, Rana Zaini Fathiyana <sup>#3</sup>

*#1,2 Pusat Jaringan Komunikasi, Badan Meteorologi Klimatologi dan Geofisika*

*#3 Program Studi Teknologi Informasi, Institut Teknologi Telkom Jakarta*

*#1,2 Jln. Angkasa 1 No.2, Kemayoran Jakarta Pusat*

*#3 Kawasan Pendidikan Telkom Jakarta, Jln. Daan Mogot Km 11*

<sup>1</sup> dinda.jaelani@bmgk.go.id

<sup>3</sup> rzainifathiyana@gmail.com

## **Abstrak**

Sistem intelijen dapat didefinisikan sebagai sistem yang menggabungkan intelijen ke dalam aplikasi yang ditangani oleh mesin. Sistem cerdas melakukan pencarian dan optimisasi bersama dengan kemampuan belajar. Sistem cerdas juga melakukan tugas otomatis yang kompleks yang tidak mungkin dilakukan dengan paradigma komputasi tradisional. Karena itu mereka dapat digunakan untuk memprediksi data meteorologi yaitu prediksi cuaca. Prediksi cuaca adalah aplikasi vital dalam meteorologi dan telah menjadi salah satu masalah yang paling menantang secara ilmiah dan teknologi di seluruh dunia pada abad terakhir. Memprediksi cuaca adalah penting untuk membantu mempersiapkan yang terbaik dan yang terburuk dari iklim. Prediksi Cuaca Akurat adalah salah satu yang paling banyak masalah yang menantang di seluruh dunia. Banyak prediksi cuaca seperti prediksi curah hujan, prediksi badai petir, prediksi kondisi cloud adalah tantangan utama untuk penelitian atmosfer. Makalah ini mengkaji penelitian yang telah dilakukan dalam prediksi cuaca dan parameter yang menyusunnya dengan bertujuan untuk menyelidiki teknik sistem cerdas digunakan serta menemukan kekuatan dan kelemahan dari setiap teknik pembelajaran dari sistem yang mungkin dapat diterapkan. Pada makalah ini menggunakan metodologi tinjauan pustaka sistematis (SLR) terkait metode prediksi cuaca yang memanfaatkan teknik pembelajaran mesin. Data yang digunakan diambil dari berbagai sumber literatur publikasi internasional. Literatur publikasi mengenai prediksi cuaca yang digunakan sebanyak 14 publikasi yang relevan dengan kombinasi kata kunci yang telah ditentukan. Kemudian dari publikasi-publikasi tersebut akan diidentifikasi dan dianalisis kekuatan dan kelemahan dari teknik pembelajaran sistem intelijen yang banyak digunakan dalam prediksi cuaca.

**Keywords:** sistem cerdas, prediksi cuaca, tinjauan pustaka sistematis

## I. PENDAHULUAN

Sistem cerdas atau *Intelligent System* sebagai paradigma terakhir dari manajemen buatan dari berbagai jenis alam, yang dapat didaftarkan sendiri dan dapat beradaptasi objek pertama mendefinisikan berdasarkan teori sistem fungsional yang menggambarkan sistem organisasi fungsi bentuk kehidupan yang memiliki hukum objektif tentang kegunaan dari efek akhir dan hasil [1].

Konsepsi sistem intelijen didefinisikan pada akhir 90-an abad XX. Ini adalah seperangkat sarana teknis dikombinasikan dengan proses informasi dan berinteraksi dengan satu orang (atau sekelompok orang) atau bekerja secara mandiri, yang mampu membuat keputusan berdasarkan informasi, pengetahuan dan insentif untuk menghasilkan tujuan dan menemukan cara rasional untuk mencapai tujuan tertentu [1]. Banyak jenis sistem intelijen seperti *fuzzy logic*, algoritma genetik, pembelajaran mesin, *neural networks*, *pattern recognition*, dan sebagainya yang sangat bermanfaat dalam aplikasi sains dan teknologi saat ini.

Prediksi adalah yang paling penting teknik penambangan data yang menggunakan seperangkat praklasifikasi contoh untuk mengembangkan model yang dapat mengklasifikasikan data dan menemukan hubungan antara data independen dan dependen [2].

Prediksi cuaca adalah aplikasi sains dan teknologi untuk memprediksi keadaan atmosfer untuk lokasi tertentu. Hal ini menjadi semakin vital bagi para ilmuwan, petani, ketahanan pangan global, manajemen bencana dan terkait organisasi untuk memahami fenomena alam untuk merencanakan dan mempersiapkan mitigasi selanjutnya. Variabel cuaca yang mempengaruhi kondisi terus menerus bervariasi dengan waktu, model prediksi dapat dikembangkan baik secara statistik atau dengan menggunakan beberapa cara lain seperti *decision tree*, *artificial neural network*, regresi, teknik pengelompokan data mining (*clustering*).

Dari banyaknya metode yang diusulkan untuk prediksi cuaca, maka pada makalah ini akan dilakukan tinjauan pustaka sistematis/ *systematic literature review* (SLR) dan akan diuraikan analisis komprehensif tentang teknik pembelajaran mesin untuk prediksi cuaca ataupun prediksi parameter cuaca khususnya prediksi suhu.

Penulisan makalah ini diatur sebagai berikut, pada Bagian II disajikan landasan teori yang berhubungan dengan penelitian, metodologi penelitian dituliskan pada Bagian III, hasil dan analisa penelitian dituliskan pada Bagian IV, terakhir di Bagian V akan disajikan kesimpulan dari penelitian.

## II. STUDI LITERATUR

P. Hemalatha [3] menerapkan metode penambangan data untuk memandu jalur kapal selama berlayar. Penentuan *Global Positioning System* digunakan untuk mengidentifikasi area di mana kapal berada sedang menavigasi. Atribut data cuaca termasuk iklim, kelembaban, suhu, badai. Laporan cuaca dari area yang dilacak dibandingkan dengan database yang ada. Dataset dianalisis disediakan untuk algoritma *decision tree*, C4.5 dan ID3. Keputusan yang didapat tentang cuaca kondisi diinstruksikan ke kapal dan jalurnya dipilih demikian. Kerja sama yang erat antara statistik dan komunitas komputasi memberikan sinergi dalam analisis data. Beberapa atribut kontinu perlu diubah karena ID3 tidak bisa langsung berurusan dengan rentang kontinu. Penelitian sebelumnya mengenai prediksi cuaca dengan menggunakan algoritma *decision tree* CART telah dilakukan oleh E. G. Petre (2009). Dalam penelitiannya data yang dikumpulkan terdaftar di Hong Kong. Data digunakan untuk membuat dataset termasuk parameter tahun, bulan, tekanan rata-rata, relative kelembaban, kuantitas awan, curah hujan dan rata-rata suhu.

S. Kannan dan S. Ghosh [5] berkontribusi terhadap pengembangan metodologi untuk memprediksi keadaan curah hujan di lokal atau regional skala untuk wilayah sungai dari data klimatologis skala besar. Sebuah model berdasarkan pada teknik pengelompokan K-mean digabungkan dengan algoritma *decision tree*, CART, digunakan untuk generasi curah hujan menyatakan dari variabel atmosfer skala besar di sungai baskom. Keadaan curah hujan harian berasal dari sejarah harian data curah hujan multi-situs dengan menggunakan K-mean clustering. Berbagai langkah-langkah validitas cluster diterapkan untuk data curah hujan yang diamati

untuk mendapatkan jumlah cluster yang optimal. CART digunakan untuk melatih data keadaan curah hujan harian dari wilayah sungai selama 33 tahun. Metodologi ini diuji untuk Sungai Mahanadi di India.

Perubahan itu diperkirakan terjadi di wilayah sungai karena pemanasan global diberikan oleh perbandingan jumlah hari jatuh di bawah negara hujan yang berbeda untuk periode yang diamati dan prediksi masa depan. Algoritma CART terbukti bagus di memprediksi kondisi curah hujan harian di wilayah sungai menggunakan *down scaling* statistik.

Soo-Yeon Ji et al. [6] memperkirakan curah hujan setiap jam di setiap waktu wilayah geografis secara efisien. Peluang hujan adalah yang pertama ditentukan. Maka hanya jika ada kemungkinan curah hujan, maka Prediksi curah hujan setiap jam dilakukan. Meskipun cukup banyak metodologi telah diperkenalkan untuk memprediksi prediksi setiap jam, kebanyakan dari mereka memiliki keterbatasan kinerja karena adanya berbagai variasi data dan jumlah terbatas data. CART dan C4.5 digunakan untuk memberikan hasil, yang dapat memberikan pola tersembunyi dan penting dengan transparan alasan. Sekitar 18 variabel digunakan dari stasiun cuaca. Untuk tujuan validasi, metode validasi silang 10 kali lipat adalah dilakukan. CART memberikan kinerja yang sedikit lebih baik daripada C4.5. Mempertimbangkan peluang, hanya sejumlah kecil contoh pergi ke prediksi yang membuatnya sulit untuk diprediksi.

F. Oliya dan A. B. Adeyemo [7] menyelidiki penggunaan data teknik penambangan dalam memprediksi suhu maksimum, curah hujan, penguapan dan kecepatan angin. *Decision tree* dengan algoritma C4.5 dan jaringan saraf tiruan digunakan untuk prediksi. Data meteorologi dikumpulkan antara tahun 2000 dan 2009 dari kota Ibadan, Nigeria. Model data untuk data meteorologi dikembangkan dan digunakan untuk melatih algoritma pengklasifikasi. Kinerja setiap algoritma adalah dibandingkan dengan metrik kinerja standar dan algoritma dengan hasil terbaik digunakan untuk menghasilkan klasifikasi aturan untuk variabel cuaca rata-rata. Saraf prediktif model jaringan juga dikembangkan untuk prediksi cuaca dan hasilnya dibandingkan dengan data cuaca aktual untuk periode yang diperkirakan. Hasilnya menunjukkan bahwa pelatihan yang diberikan cukup data, teknik penambangan data dapat digunakan secara efisien untuk cuaca studi prediksi dan perubahan iklim.

Subana Shanmuganathan dan Philip Sallis [8] memeriksa penggunaan metode penambangan data untuk mencari pola dalam kondisi cuaca ad hoc, seperti waktu hari, bulan tahun, arah angin, kecepatan, dan tingkat keparahan menggunakan set data dari Lokasi tunggal. Data cuaca historis, antara 2008 dan 2008 2012 digunakan dari perangkat telemetri yang dipasang di kebun anggur di utara Selandia Baru. Terlihat bahwa menggunakan data mining teknik dan kondisi cuaca setempat dicatat tidak beraturan Interval dapat menghasilkan pengetahuan baru yang berkaitan dengan hembusan angin pola pengambilan keputusan manajemen kebun anggur. Dari penyimpanan data, contoh yang berkaitan dengan kebun anggur Sungai Kumeu diekstraksi untuk periode empat tahun (2008-2012). Data dikumpulkan dibersihkan untuk menghapus semua bacaan yang ada di luar Kumeu merekam bacaan. Contoh terakhir 86.418 dan mereka distribusi selama 12 bulan disajikan. *Decision tree* dengan algoritma yang digunakan adalah C5, Quest, CRT dan CHAID. SOM digunakan untuk tujuan pengelompokan. JST yang diawasi multilayer adalah digunakan untuk memprediksi hembusan angin. Teknik dan data mining metode statistik dijalankan menggunakan SPSS. Ini menyediakan alat yang bagus untuk menganalisis dataset adhoc.

Kavita Pabreja [9] menunjukkan derivasi dari sub-grid sistem cuaca skala dari produk keluaran model NWP menggunakan teknik penambangan data yang tidak mungkin melalui normal Teknik MOS. Teknik data mining, clustering, kapan diterapkan pada divergensi dan kelembaban relatif dapat memberikan indikasi awal pembentukan *cloudburst*. K berarti pengelompokan digunakan untuk data dua hari dari kasus nyata *cloudburst*. Sebuah upaya dilakukan untuk menyediakan tepat waktu dan dapat ditindaklanjuti informasi peristiwa ini menggunakan teknik penambangan data di Indonesia suplemen dengan model NWP. Satu kekurangan ditemukan itu itu tidak dapat digunakan untuk prediksi jangka panjang.

M. A. Kalyankar dan S. J. Alaspurkar [10] menggunakan data mining teknik untuk memperoleh data cuaca dan menemukan pola yang tersembunyi di dalam dataset besar untuk mentransfer yang diambil informasi menjadi pengetahuan yang dapat digunakan untuk klasifikasi dan prediksi kondisi cuaca. Proses penambangan data diterapkan untuk mengekstrak pengetahuan dari dataset cuaca kota Gaza. Ini pengetahuan dapat digunakan untuk memperoleh prediksi dan dukungan yang bermanfaat proses pengambilan keputusan. Metode

penambahan data dinamis diperlukan untuk membangun, yang dapat belajar secara dinamis untuk mencocokkan sifat cuaca yang berubah dengan cepat dan kejadian tiba-tiba.

S. S Badhiye et al. [11] menggunakan teknik pengelompokan dengan metode K Nearest Neighbor untuk menemukan pola tersembunyi di dalamnya dataset besar yang terkait dengan cuaca sehingga dapat mentransfer yang diambil informasi menjadi pengetahuan yang dapat digunakan untuk klasifikasi dan prediksi kondisi iklim. Suhu dan kelembabannya diperoleh untuk interval waktu tertentu. Prediksi tinggi akurasi diperoleh untuk suhu dan kelembaban. Itu perangkat lunak dapat tertanam dengan sistem data logger untuk analisis dan prediksi parameter di daerah terpencil.

Pinky Saikia Dutta dan Hitesh Tahbilder [12] memperkirakan Curah Hujan Assam bulanan dengan menggunakan teknik data mining. Teknik statistik tradisional -Beberapa Regresi Linier digunakan. Data termasuk periode Enam tahun antara 2007 dan 2012 yang dikumpulkan secara lokal dari Pusat Meteorologi Regional, Guwahati, Assam, India. Data dibagi menjadi empat bulan untuk setiap musim. Parameter yang dipilih untuk model adalah suhu minimum, suhu maksimum, tekanan permukaan laut rata-rata, kecepatan angin dan curah hujan. Kinerja model ini diukur dalam penyesuaian R-kuadrat yang diimplementasikan dalam C#. Beberapa parameter seperti arah angin tidak dimasukkan karena kendala pada pengumpulan data yang dapat memberikan hasil yang lebih akurat. Akurasi yang dapat diterima diberikan oleh model prediksi berdasarkan regresi linier berganda.

Neha Khandelwal dan Ruchi Davey [13] memperkirakan curah hujan satu tahun dengan menggunakan 4 faktor iklim suhu yang berbeda, kelembaban, tekanan dan permukaan laut dan dengan demikian menggunakan dataset untuk menghitung kemungkinan kekeringan di Rajasthan. Faktor tertentu diekstraksi menggunakan teknik penambahan data. Kemudian analisis korelasi diterapkan pada dataset dan korelasi adalah ditemukan dalam faktor-faktor. Faktor-faktor dengan korelasi positif adalah dipilih dan digunakan untuk analisis regresi. MLR digunakan untuk analisis regresi untuk memprediksi curah hujan. Lalu statistikanalisis diterapkan pada data tersebut untuk menemukan kemungkinan kekeringan. Untuk kemungkinan standar deviasi kekeringan, varians koefisien, indeks kekeringan dan persepsi kekeringan digunakan.

A.R.W.M.M.S.C.B. Amarakoon [14] mengusulkan sistem itu menggunakan data cuaca historis dan menerapkan penambahan data algoritma "*K-Nearest Neighbor (KNN)*" untuk klasifikasi data historis ini menjadi rentang waktu tertentu. K terdekat rentang waktu selanjutnya diambil untuk memprediksi cuaca Sri Lanka. Data cuaca harian dikumpulkan untuk lengkap satu tahun. Ini menghasilkan hasil yang akurat dalam waktu yang wajar selama berbulan-bulan sebelumnya. Disimpulkan bahwa KNN bermanfaat untuk data dinamis, data yang berubah atau diperbarui dengan cepat dan memberikan kinerja yang lebih baik dibandingkan dengan yang lain teknik. Integrasi teknik pemilihan fitur bahkan bisa memberikan hasil yang lebih akurat.

Gaurav J. Sawale dan Sunil R. Gupta [15] mengusulkan buatan metode jaringan saraf untuk prediksi cuaca di masa depan di lokasi tertentu. *Back Propagation Neural Network* digunakan untuk pemodelan awal. Kemudian Hopfield Networks diberi makan dengan hasil yang dikeluarkan oleh model BPN. Atribut termasuk suhu, kelembaban dan kecepatan angin. Tiga tahun data cuaca dikumpulkan terdiri dari 15000 kejadian. Kesalahan prediksi sangat kurang dan proses belajar cepat. Ini dapat dianggap sebagai alternatif dari tradisional pendekatan meteorologi. Kedua algoritma digabungkan secara efektif. Hal ini dapat menentukan hubungan non-linear itu ada di antara atribut data historis dan memprediksi cuaca di masa depan.

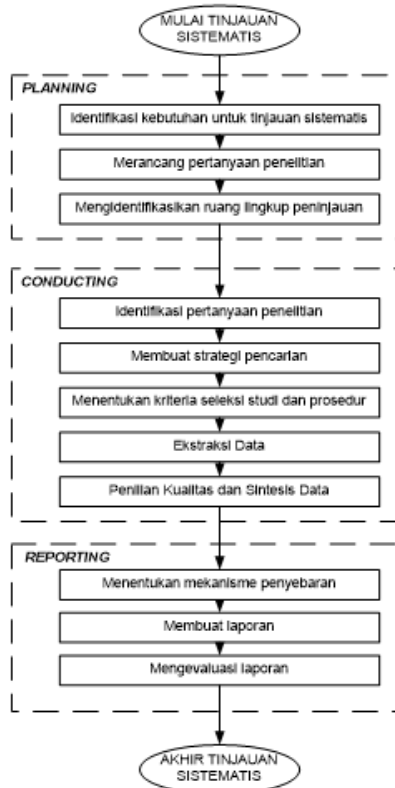
Alotaibi et al. [16] Memprediksi dari pola curah hujan di daerah yang langka air sangat penting untuk manajemen sumber daya air yang efektif. Model berbasis data termasuk jaringan saraf tiruan (JST) dan neuro adaptif sistem inferensi fuzzy (ANFIS) dapat digunakan untuk memprediksi perubahan curah hujan dan suhu jangka panjang di masa depan. Dalam penelitian ini data iklim bulanan selama tiga puluh tahun digunakan untuk analisis tren diuji menggunakan Mann-Kendall untuk mensimulasikan perubahan suhu dan curah hujan menggunakan tiga GCM (yaitu, HADCM3, INCM3, dan MPEH5) untuk skenario emisi A1B, A2, dan B1 serta dua model berbasis data (JST: feed-forward-multilayer, perceptron dan ANFIS). Hasil menunjukkan bahwa JST dapat mengungguli ANFIS untuk memprediksi suhu dan curah hujan jangka panjang di masa depan dengan akurasi yang dapat diterima. Secara keseluruhan, prediksi masa depan menunjukkan bahwa suhu Wilayah

Qassim akan meningkat dengan pola tertentu dari 2011 hingga 2099, sedangkan perubahan curah hujan akan berbeda pada berbagai rentang masa depan.

Mohideen et al. [17] Prediksi curah hujan sangat membantu untuk sektor pertanian. Prediksi awal kekeringan dicapai melalui data deret waktu. Untuk tepatnya prediksi, teknik Jaringan Syaraf Tiruan (JST) digunakan. Dataset hujan diuji menggunakan *Feed Forward Neural Network* (FFNN). Kinerja model ini dievaluasi menggunakan *Mean Square Error* (MSE) dan *Magnitude of Relative Error* (MRE). Kinerja yang lebih baik dicapai jika dibandingkan dengan teknik penambahan data lainnya

### III. METODOLOGI PENELITIAN

Dalam melakukan studi kepustakaan yang mengulas mengenai topik pengembangan sistem intelijen untuk prediksi cuaca khususnya dalam pengukuran suhu, dipergunakan sebuah pendekatan secara sistematis berupa *systematic literature review* (SLR). SLR sekarang ini banyak dipergunakan sebagai metode pengulasan dalam penelitian di bidang rekayasa perangkat lunak. SLR didefinisikan sebagai sebuah proses yang di dalamnya melakukan identifikasi, penilaian, dan interpretasi seluruh barang bukti penelitian yang terseedia dengan tujuan untuk menjawab sejumlah pertanyaan penelitian. Studi kepustakaan yang dilakukan untuk penelitian ini mengadopsi pada sumber atau panduan SLR yang dikemukakan oleh Kitchenham and Charters (2007).



Gambar 1. Langkah-langkah tinjauan sistematis [18]

Metodologi yang digunakan terdiri dari tiga fase utama yaitu: perencanaan (*planning*), pengerjaan (*conducting*), dan pelaporan (*reporting*) [18]. Dimana pada setiap fasenya memiliki langkah-langkah yang lebih

spesifik, yang akan dijelaskan lebih mendalam pada bagian ini. Gambar 1 merangkum tahapan pelaksanaan tinjauan pustaka sistematis yang digunakan pada penelitian ini:

Fase awal berupa perencanaan (*planning*) yang dibagi ke dalam tiga tahap yaitu: (a) identifikasi kebutuhan untuk kajian sistematis, (b) merancang pertanyaan penelitian, dan (c) mendefinisikan ruang lingkup peninjauan. Pada fase ke-dua dilakukan perancangan (*conducting*) yang terdiri dari: (a) identifikasi pertanyaan penelitian, (b) membuat strategi pencarian, (c) menentukan kriteria seleksi studi dan prosedur, (d) ekstraksi data, (e) penilaian kualitas dan sintesis data. Sedangkan pada fase pelaporan (*reporting*) terdiri dari: (a) menentukan mekanisme penyebaran, (b) membuat laporan sesuai format, dan (c) mengevaluasi laporan. Fase yang paling krusial dan penting adalah pada fase *conducting*, karena merupakan tahapan pelaksanaan. Pada tahap ke-tiga dilakukan evaluasi secara iteratif terhadap protokol pengulasan literatur yang dipergunakan selama fase pengerjaan (*conducting*) dijalankan.

#### A. Identifikasi Pertanyaan Penelitian

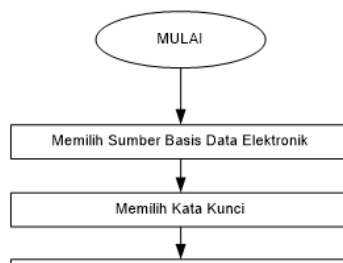
Dalam SLR pertanyaan penelitian diistilahkan sebagai *research question* (RQ) yang memainkan peran penting dalam menentukan strategi pencarian, ekstraksi data, dan kemudian tahap analisis, Identifikasi pertanyaan penelitian bertujuan untuk menemukan informasi sebanyak mungkin dengan menggunakan strategi pencarian yang sesuai, selain itu pada langkah ini juga bertujuan untuk menghindari bias publikasi. Penelitian ini bertujuan untuk meninjau studi yang telah dilakukan dalam pengembangan sistem intelijen yang digunakan untuk memprediksi suhu yang kemudian akan dianalisis kekuatan dan kelemahan dari setiap teknik pembelajaran mesin yang digunakan. Untuk mempermudah pencarian dan membatasi ruang lingkup rencana penelitian maka dibuat pertanyaan penelitian yang didefinisikan pada Tabel I.

#### B. Strategi Pencarian

Menurut Aline dkk [19], strategi pencarian dimulai dari *review question* dan kerangka konseptual yang dimaksudkan untuk dapat mengetahui apa yang harus dicari, bagaimana meminimalkan bias penelitian, mempertimbangkan penelitian yang relevan, dan seberapa luas pencariannya. Selanjutnya langkah awal yang ditempuh adalah dengan menentukan istilah pencarian yakni apa yang harus dicari. Menentukan istilah pencarian saling terkait dan seiring dengan pemilihan sumber data pencarian, serta mempertimbangkan kriteria inklusi dan pengecualian.

Sumber data pencarian yang digunakan menggunakan perpustakaan elektronik IEEE Xplore (<https://ieeexplore.ieee.org/>) yang diakses mulai dari tanggal 1Juni 2020. Pencarian online terkait dengan konferensi dan makalah jurnal yang akan dieksplorasi adalah publikasi dalam satu dekade dimulai dari tahun 2010. Adapun alur proses strategi pencarian ditunjukkan pada Gambar 2. Strategi pencarian dilakukan dengan mencari kata kunci yang relevan dengan penelitian. Istilah pencarian dapat dibangun oleh salah satu operator Boolean “AND” dan “OR” yang digunakan untuk mengambil kombinasi istilah tertentu untuk pencarian pada bagian kata judul, abstrak, dan kata kunci yang digunakan rangkaian kata sebagai berikut:

((“temperature”) OR (“weather”) AND (“prediction”) AND (“neural network”) OR (“data mining”))



Gambar 2. Strategi Pencarian

TABEL I. PERTANYAAN PENELITIAN

ID	Pertanyaan Penelitian (RQ)	Tujuan
RQ1	Topik seperti apa yang dipilih peneliti sebagai objek parameter cuaca dalam prediksi dengan menggunakan sistem intelijen?	Identifikasi jenis topik yang dipilih oleh peneliti dalam melakukan prediksi dengan menggunakan sistem intelijen
RQ2	Teknik pembelajaran mesin apa yang digunakan atau diusulkan oleh peneliti dalam mengembangkan sistem intelijen untuk prediksi suhu?	Untuk mendapatkan pengetahuan yang mendalam, mengidentifikasi, dan mengelompokan metode sistem intelijen / teknik pembelajaran mesin yang digunakan untuk memprediksi cuaca.
RQ3	Berapa tingkat akurasi yang didapat dari teknik pembelajaran mesin yang digunakan untuk prediksi suhu?	Untuk mengetahui besarnya keakurasian teknik pembelajaran mesin dalam memprediksi suhu.
RQ4	Kelebihan dan kekurangan dari tiap teknik pembelajaran mesin dalam prediksi cuaca khususnya pada parameter suhu?	Untuk mengetahui hasil perbandingan antara teknik pembelajaran mesin yang digunakan untuk prediksi suhu

TABEL II. PROPERTI YANG DIEKSTRAKSI

	Properti	Pertanyaan Penelitian
P1	Topik penelitian	RQ1
P2	Metode teknik pembelajaran mesin untuk prediksi suhu	RQ2
P3	Indikator Keakuratan	RQ3
P4	Karakteristik teknik pembelajaran mesin untuk prediksi suhu	RQ4

### C. Kriteria Seleksi Studi

Untuk mempermudah mendapatkan publikasi yang sesuai dengan penelitian maka langkah selanjutnya adalah membuat filter dalam pemilihan dan penolakan sesuai kriteria yang telah didefinisikan. Adapun kriteria utama guna menunjang penelitian ini adalah terkait sistem intelijen yang digunakan untuk memprediksi cuaca

dan kemudian akan ditinjau sistem intelijen atau teknik pembelajaran mana yang paling sering digunakan serta dilihat dari keakuratan dari tiap teknik pembelajaran mesinnya. Maka penyaringan dilakukan dengan mengeliminasi studi prediksi cuaca yang tidak ditunjukkan untuk pengukuran suhu suatu daerah. Selanjutnya studi yang tidak menggunakan metode prediksi dengan teknik pembelajaran mesin maka tidak akan dipertimbangkan. Kemudian, tahun publikasi yang dipertimbangkan dimulai dari tahun 2010, maka publikasi yang penelitiannya dilakukan sebelum tahun 2010 akan dieliminasi. Terakhir, publikasi yang tidak dituliskan dalam bahasa Inggris atau bahasa Indonesia akan dikecualikan.

Pencarian diawali dengan memasukan kata kunci prediksi temperature / cuaca, dari IEEE Xplore didapat 634 publikasi. Selanjutnya dieliminasi menjadi 47 publikasi dengan menambahkan penyaringan publikasi hanya yang menggunakan teknik pembelajaran mesin atau neural network. Setelah menerapkan kriteria seleksi studi Dari 47 publikasi terdapat 33 publikasi yang tidak relevan, walaupun tidak terdapat publikasi yang tidak sesuai dengan kriteria studi baik dari kelengkapan teks maupun bahasa yang digunakan. Sehingga hasil akhir didapat 14 publikasi yang kemudian akan dibaca secara keseluruhan dan dianalisis untuk kemudian dapat menjawab pertanyaan penelitian yang telah dibuat.

#### D. Ekstraksi Data

Dari 14 publikasi yang dijadikan studi utama selanjutnya dilakukan ekstraksi data dengan iterasi hingga didapat data untuk menjawab pertanyaan penelitian yang diajukan. Adapun properti untuk menunjukkan pemetaan untuk masing-masing pertanyaan penelitian tercantum pada Tabel II. Kelima properti (P1-P5) merupakan karakteristik yang diharapkan dari 14 publikasi yang diekstraksi.

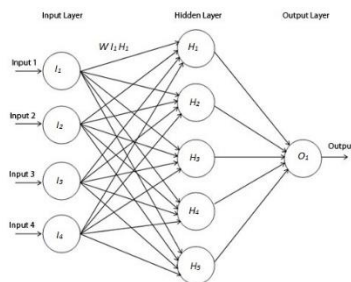
##### 1. Topik Penelitian (P1)

Dari hasil pencarian ditemukan bahwa publikasi yang didapat dan dijadikan bahan utama dalam SLR ini tidak hanya membahas prediksi suhu saja. Terdapat beberapa parameter cuaca lainnya seperti kelembaban, tekanan, arah angin, curah hujan, dan lain-lain. Dari sejumlah publikasi mengenai prediksi cuaca selanjutnya peneliti memklasifikasikan domain apa saja dan berapa presentase dari setiap parameter penelitiannya.

##### 2. Metode teknik pembelajaran mesin prediksi suhu (P2)

- *Artificial Neural Network*

Metode yang mengadopsi mekanisme berpikir sebuah sistem yang menyerupai otak manusia. Neuron dan sinapsis diwakili oleh grafik simpul dan tepi. Seperti yang ditunjukkan oleh gambar 3, dalam jaringan bagian lapisan simpul terdapat pemodelan input. Kemudian distribusi bobot diterapkan ke setiap koneksi dalam grafik.



Gambar 3. Diagram Artificial Neural Network dengan satu Hidden Layer

Setiap node menganggap input sebagai fungsi simpul; dan untuk simpul yang terhubung ke lapisan sebelumnya. Untuk setiap neuron, sinyal yang diterima adalah seperti yang diberikan di bawah ini:

$$u_j = \sum W_{ij} * X_i \quad (1)$$



Dimana,  $W_{ij}$  adalah bobot relasi antara neuron  $i$  dan  $j$  dan  $X_i$  adalah input dari  $i$ . Jika hasilnya lebih besar dari nilai predikat neuron, maka itu menjadi input untuk layer konsekuen.

- *Decision Tree*

*Decision tree* merupakan metode yang berusaha menemukan fungsi-fungsi pendekatan yang bernilai diskrit dan tahan terhadap data-data yang memiliki kesalahan (*noise*). Memiliki algoritma pada metode ini digunakan untuk kasus klasifikasi data. Sebuah *decision tree* terdiri dari sebuah node paling tinggi disebut *root*, *internal node* dan *leaf*.

- *K-Nearest Neighbor Algorithm*

K-Nearest Neighbor (k-NN) merupakan teknik klasifikasi data berdasarkan kedekatan jarak (lokasi) suatu data terhadap data lainnya [1]. Jarak yang digunakan adalah *Euclidean Distance*. Jarak Euclidean (*Euclidean Distance*) merupakan perhitungan yang paling umum digunakan pada data numerik. Rumus *Euclidean Distance*:

$$D_{ij} = \sqrt{\sum_{k=1}^n (X_{ij} - X_{jk})^2} \quad k = 1, 2, \dots, n \quad (2)$$

Keterangan:

- $i$  : Sampel Data
- $j$  : Data Uji / Testing
- $k$  : Variabel Data
- $D$  : Jarak

Untuk setiap titik data dalam dataset, jarak Euclidean antara titik data input dan titik saat ini. Jarak ini diurutkan secara meningkat dan  $k$  item dengan jarak terendah ke titik data input yang dipilih

- *Linier Regression*

Metode ini digunakan untuk klasifikasi biner, yang berdasar pada metode statistik. Menggunakan model linier sehingga digunakan untuk melakukan regresi pada sekelompok variabel. Teknik ini sering digunakan untuk memprediksi pola dalam data dengan atribut yang tidak ambigu atau numerik. Serta untuk menghitung probabilitas menggunakan serangkaian vektor input dan variabel respons dependen, menggunakan logaritma. Probabilitas terletak di antara kelas tertentu. Untuk klasifikasi biner, variabel respons yang diberikan di bawah ini:

$$Y_i = \begin{cases} 0 \\ 1 \end{cases} \quad (3)$$

Kemudian, rumus untuk menghitung bahwa sampel  $x_i$  termasuk dalam kelas satu ditunjukkan oleh formula matematika sebagai berikut. Dimana  $W_0$  dan  $W$  adalah patameter standarisasi regresi.

$$P(y_i = 1 | X_i) = \frac{\exp(W_0 + W^T X_i)}{1 + \exp(W_0 + W^T X_i)} \quad (4)$$

### 3. Indikator Keakuratan (P3)

Dari publikasi yang dianalisis, Peneliti menggunakan indikator ketelitian dari setiap teknik pembelajaran mesin yang digunakan. Empat metrik dasar yang digunakan dalam mengevaluasi percobaan, yaitu *True Positive* (TPR), *True Negative* (TNR), *False Positive* (FPR) dan *False Negative* (FNR).

		Positif	Negatif
Nilai Aktual	Positif	TP	FN
	Negatif	FP	TN
		Prediksi Hasil	

Gambar 4. Confusion Matrix

Adapun parameter yang akan dianalisis adalah parameter *accuracy*. *Accuracy* merupakan rasio prediksi benar (positif dan negatif) dengan keseluruhan data. Dengan kata lain, *accuracy* merupakan tingkat kedekatan nilai prediksi dengan nilai aktual (sebenarnya). Nilai *accuracy* dapat diperoleh dengan persamaan:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

#### 4. Karakteristik teknik pembelajaran mesin untuk prediksi suhu (P4)

Peneliti melakukan perbandingan dari teknik pembelajaran mesin yang digunakan dalam prediksi cuaca khususnya parameter suhu. Selanjutnya dianalisis apa saja kekurangan dan kelebihan dari masing-masing teknik pembelajaran mesin yang digunakan.

#### E. Penilaian Kualitas Studi dan Sintesis Data

Penilaian kualitas studi merupakan tahapan yang dapat memberikan bantuan dalam interpretasi data yang berhasil diekstraksi sebelumnya. Kualitas data akan bergantung pada bagaimana data dikumpulkan sebagai barang bukti untuk menjawab RQ. Data sederhana akan mengakibatkan kurangnya barang bukti yang dikumpulkan, namun apabila data hasil ekstraksi didapatkan dalam jumlah besar, maka kualitas barang bukti akan mampu menjawab RQ yang sudah ditentukan di tahap sebelumnya.

Sintesis data dapat diartikan sebagai interpretasi visualisasi dari sebuah data yang semula masih disajikan mentah. Sintesis data dilakukan untuk memperkuat analisa dan presentasi bagi peneliti dalam mengolah dan menyajikan bagaimana sistem intelijen untuk prediksi cuaca. Sintesis data dilakukan dengan alat bantu visualisasi, seperti diagram batang, diagram bulat, ataupun infografis lainnya

## IV. HASIL DAN DISKUSI

Pada bab ini menyajikan hasil penelitian dari tahapan yang sudah dilakukan menggunakan metode tinjauan pustaka sistematis (SLR).

### RQ1 – Topik Penelitian

Analisis dari literatur yang menjadi bahan utama SLR menunjukkan adanya beberapa topik penelitian selain prediksi suhu dalam pengembangan sistem intelijen. Topik lain ini dimaksudkan kepada parameter cuaca lainnya, seperti kelembaban, tekanan, arah angin, curah hujan, dan lain-lain. Tabel III merupakan tabel distribusi topik penelitian parameter cuaca dalam sistem prediksi dengan sistem intelijen.

**B. RQ2 – Teknik Pembelajaran Mesin digunakan pada Penelitian Prediksi Suhu**

Pada bagian ini disajikan hasil analisis dari teknik pembelajaran mesin yang digunakan dalam penelitian prediksi suhu. Bagian analisis ini dimaksudkan untuk menjawab RQ2. Teknik pembelajaran mesin yang disajikan hanya yang digunakan untuk tinjauan sistematis (SLR) ini. Akan diidentifikasi yang paling banyak digunakan untuk domain prediksi suhu yang berkaitan dengan prediksi suhu. Seperti pada tabel IV distribusi teknik pembelajaran mesin yang digunakan pada prediksi suhu.

Tabel IV merupakan hasil pemetaan dan analisis terhadap 14 publikasi yang dijadikan studi utama. Dapat diketahui bahwa teknik pembelajaran mesin yang paling banyak digunakan untuk prediksi suhu adalah dengan menggunakan Artificial Neural Network (ANN) yaitu sebanyak 6 publikasi yang mengusulkan menggunakan metode ANN dalam prediksi suhu. Kemudian dalam sejumlah 5 publikasi menggunakan metode Decision Tree. Dan metode lainnya seperti Clustering (K-NN) dan Regresion pun digunakan sebagai sistem intelijen untuk prediksi suhu. Untuk selanjutnya metode teknik pembelajaran tersebut akan dianalisis karakteristiknya, apa saja yang menjadi kekuatan dan kelemahan dalam metodenya. Ini dimaksudkan untuk menjawab pertanyaan penelitian pada RQ4.

**C. RQ3 – Tingkat Keakuratan dari Teknik Pembelajaran Mesin untuk Prediksi Suhu.**

Dari 14 publikasi yang dijadikan studi utama dipilih 4 publikasi yang mewakili setiap metode teknik pembelajaran yang ditinjau. Hasil pemetaan analisis keakuratan dari teknik pembelajaran mesin untuk prediksi suhu seperti pada tabel V.

**D. RQ4 - Kekuatan dan Kelemahan dari Teknik Pembelajaran Mesin pada Prediksi Suhu**

Dari semua teknik pembelajaran mesin yang untuk tinjauan sistematis (SLR) ini, tidak terdapat satu pun teknik pembelajaran mesin yang dapat secara seragam mengungguli teknik lainnya pada semua data set. Tiap-tiap teknik memiliki kekuatan dan kelemahan tersendiri. Berikut bukti studi empiris dan teoritis yang disajikan dalam bentuk tabel perbandingan untuk masing-masing teknik pembelajaran mesin seperti pada tabel VI.

TABEL III. DISTRIBUSI PARAMETER CUACA DALAM SISTEM PREDIKSI INTELIGEN

Publikasi	Parameter								
	Temperature	Kelembaban	Tekanan	Curah Hujan	Arah Angin	Kecepatan Angin	Kuantitas Awan	Sea Level	Pengedapan
P. Hemalatha [3]	✓	✓	-	-	-	-	-	-	-
S Kannan, S Ghosh [5]	✓	-	✓	-	✓	✓	-	✓	-
S Yeon <i>et al</i> [6]	✓	✓	✓	-	✓	✓	-	-	-
F Olaiya, AB Adeyemo [7]	✓	-	-	✓	-	✓	-	-	✓
P Sallis, Subana S [8]	✓	✓	-	-	-	✓	-	-	-
K Pabreja [9]	✓	✓	-	-	-	-	-	-	-
M.A Kalyankar [10]	✓	✓	-	✓	-	✓	-	-	-
S Badhiye <i>et al</i> [11]	✓	✓	-	-	-	-	-	-	-

Publikasi	Parameter								
	Temperature	Kelembaban	Tekanan	Curah Hujan	Arah Angin	Kecepatan Angin	Kuantitas Awan	Sea Level	Pengedapan
PS Dutta [12]	✓	✓	-	✓	✓	-	-	-	-
N Khandelwal [13]	✓	✓	-	✓	-	-	-	✓	-
Amarakoon [14]	✓	✓	-	-	✓	✓	-	-	-
GJ Sawale [15]	✓	-	-	✓	-	-	-	-	-
Alotaibi [16]	-	-	-	✓	-	-	-	-	-
Mohideen <i>et al</i> [17]	-	-	-	✓	-	-	-	-	-

TABEL IV. DISTRIBUSI PUBLIKASI TEKNIK PEMBELAJARAN MESIN UNTUK PREDIKSI SUHU

No	Teknik Pembelajaran Mesin	Sumber Literatur
1	Artificial Neural Network	[7], [8], [14], [15], [16], [17]
2	Decision Tree	[3], [5], [6], [7], [8]
3	Clustering	[5], [9], [10], [11]
4	Regression	[12], [13]

TABEL V. TINGKAT ACCURACY TEKNIK PEMBELAJARAN MESIN UNTUK PREDIKSI SUHU

Teknik Pembelajaran Mesin	Tingkat Akurasi		
	Mohideen Shah <i>et al</i> , 2019 [17]	F. Olaiya and A. B. Adeyemo, 2012 [7]	PS Dutta, H Tahbilder, 2014 [12]
Artificial Neural Network	96.6 %		
Decission Tree		82%	
Regression			63%

TABEL VI. PERBANDINGAN TEKNIK PEMBELAJARAN MESIN UNTUK PREDIKSI SUHU

Parameter	Teknik Pembelajaran Mesin		
	Artificial Neural Network	Decision Tree	Clustering
Performance	Maximum	High	Moderate
Supervised Learning	Yes	Yes	No
Unsupervised Learning	Yes	No	Yes
Kecepatan Komputasi	Fast	Fastest	Slow
Cost Effectiveness	Yes	Yes	Yes
Akurasi	More	Less	More
Tingkat Kompleksitas	Complex	Less	Moderate
Fault Tolerance	Yes	Yes	Yes
Domain Area	Complex	Simple	Complex
Parallel Computing	Yes	Yes	No

Dari 14 publikasi utama yang menggunakan beberapa metode dan teknik/pendekatan berbeda-beda berdasarkan berbagai parameter, ANN muncul menjadi teknik yang paling kuat yang dapat memberikan hasil yang akurat dan tepat waktu. Ia dapat menangani nilai real dan non real yang kompleks, memiliki kecepatan komputasi terbesar, memberikan hasil yang lebih akurat, dan juga memberikan hasil yang lebih baik untuk set data multidimensi. Perkembangan penelitian machine learning dalam hal prediksi semakin berkembang setiap tahunnya dengan peningkatan jumlah penelitian yang menggunakan ANN dalam metode prediksi cuaca

maupun parameter yang berkaitan dengan cuaca atau iklim. Hal tersebut menunjukkan bahwa penggunaan ANN dalam sistem prediksi cuaca memiliki perkembangan yang baik.

## V. KESIMPULAN

Tinjauan sistematis ini bertujuan untuk mengidentifikasi dan menganalisis kecenderungan metode, dan teknik/pendekatan yang digunakan dalam penelitian di bidang pengembangan sistem intelijen untuk prediksi suhu mulai dari tahun 2010 sampai dengan tahun 2020. Berdasarkan kriteria pemilihan keputusan, didapatkan 14 penelitian terpublikasi yang berelasi dengan pengembangan sistem intelijen untuk prediksi cuaca yang didalamnya terdapat pula parameter prediksi suhu yang menjadi bahan utama analisis menggunakan skema SLR. SLR merupakan sebuah proses yang melakukan identifikasi, penilaian, dan penerjemahan seluruh data dan barang bukti penelitian yang tersedia dengan tujuan untuk menjawab sejumlah pertanyaan penelitian (RQ).

Sebanyak 4 metode dimanfaatkan sebagai metode dalam mengembangkan sistem intelijen untuk prediksi suhu. Metode yang umum digunakan adalah dengan memanfaatkan teknik pembelajaran ANN, dimana terdapat 6 publikasi menggunakannya, Decision Tree sebanyak 5 literatur, Clustering sebanyak 4 publikasi, dan 2 publikasi menggunakan teknik Regression.

Topik penelitian dari 14 publikasi yang dianalisis menunjukkan terdapat 9 parameter cuaca dalam pengembangan sistem intelijen untuk prediksi cuaca. Dari ke-14 publikasi parameter cuaca suhu ditinjau oleh 11 publikasi. Ini menunjukkan bahwa parameter suhu merupakan parameter yang diminati dan menjadi dasar dalam prediksi cuaca.

Dari 4 metode yang dimanfaatkan sebagai metode dalam mengembangkan sistem intelijen untuk prediksi suhu, kemudian ditinjau tingkat akurasi nya, hasil menunjukkan bahwa metode ANN merupakan metode dengan tingkat akurasi paling tinggi yaitu sebesar 96.6 %. Selain itu ANN juga memiliki *performance* terbaik dibanding dengan metode lainnya. Berdasarkan hal ini maka untuk penelitian selanjutnya, peneliti akan menggunakan sistem intelijen dengan teknik pembelajaran ANN.

## PENGHARGAAN

Penulis mengucapkan banyak terima kasih kepada pihak-pihak terkait yang membantu dalam mendukung pembuatan penelitian ini seperti Pusat Jaringan Komunikasi Badan Meteorologi Klimatologi dan Geofisika (BMKG) dan Institut Teknologi Telkom Jakarta.

## REFERENSI

- [1] E. Hines, *Intelligent systems: techniques and applications*. Maastricht: Shaker Publishing, 2008.
- [2] D. Chauhan dan J. Thakur, "Data Mining Techniques for Weather Prediction: A Review," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 2, no. 8, hlm. 6.
- [3] P.Hemalatha, "Implementation of Data Mining Techniques for Weather Report Guidance for Ships Using Global Positioning System," dalam *International Journal Of Computational Engineering Research*, Mar 2013, vol. Vol. 3 Issue. 3.
- [4] Elia Georgiana Petre, "Decision Tree for Weather Prediction," 2009, vol. LXI No. 1, hlm. 77–82.
- [5] S. Kannan , Subimal Ghosh, "Prediction of daily rainfall state in a river basin using statistical downscaling from GCM output," Jul 2010.
- [6] Soo-Yeon Ji , Sharad Sharma, Byunggu Yu, dan Dong Hyun Jeong, "Designing a Rule-Based Hourly Rainfall Prediction Model," dipresentasikan pada IEEE IRI 2012, Agu 2012.
- [7] Olaiya, F., & Adeyemo, A. B, "Application of Data Mining Techniques in Weather Prediction and Climate Change Studies," dalam *International Journal of Information Engineering and Electronic Business*, 2012, vol. 4(1), hlm. 51–59.

- [8] Subana Shanmuganathan and Philip Sallis, "Data Mining Methods to Generate Severe Wind Gust Models," dalam *Atmosphere 2014*, vol. 5, hlm. 60–80.
- [9] Kavita Pabreja, "Clustering technique to interpret Numerical Weather Prediction output products for forecast of Cloudburst," dalam *International Journal of Computer Science and Information Technologies (IJCSIT)*, 2012, vol. Vol. 3 (1), hlm. 2996–2999.
- [10] Meghali A. Kalyankar, S. J. Alasapurkar, "Data Mining Technique to Analyse the Metrological Data," dalam *International Journal of Advanced Research in Computer Science and Software Engineering 3(2)*, Feb 2013, hlm. 114–118.
- [11] Badhiye S. S., Dr. Chatur P. N., Wakode B. V., "Temperature and Humidity Data Analysis for Future Value Prediction using Clustering Technique: An Approach," dalam *International Journal of Emerging Technology and Advanced Engineering*, Jan 2012, vol. Volume 2, Issue 1.
- [12] Pinky Saikia Dutta, Hitesh Tahbilde, "Prediction Of Rainfall Using Data mining Technique Over Assam," dalam *Indian Journal of Computer Science and Engineering (IJCE)*, Mei 2014, vol. Vol. 5.
- [13] Neha Khandelwal, Ruchi Davey, "Climatic Assessment Of Rajasthan's Region For Drought With Concern Of Data Mining Techniques," dalam *International Journal Of Engineering Research and Applications (IJERA)*, Okt 2012, vol. Vol. 2, hlm. 1695–1697.
- [14] A.R.W.M.M.S.C.B. Amarakoon, "Effectiveness of Using Data Mining for Predicting Climate Change in Sri Lanka," dipresentasikan pada 2014 9th Latin American Web Congress, Ouro Preto, Brazil, Okt 2014.
- [15] Gaurav J. Sawale, Dr. Sunil R. Gupta, "Use of Artificial Neural Network in Data Mining For Weather Forecasting," dalam *International Journal Of Computer Science And Applications*, Apr 2013, vol. Vol. 6.
- [16] K. Alotaibi, A. Ghumman, H. Haider, Y. Ghazaw, dan Md. Shafiquzzaman, "Future Predictions of Rainfall and Temperature Using GCM and ANN for Arid Regions: A Case Study for the Qassim Region, Saudi Arabia," *Water*, vol. 10, no. 9, hlm. 1260, Sep 2018, doi: 10.3390/w10091260.
- [17] "Soft Computing Research For Weather Prediction Using Multilayer Architecture," *ijeat*, vol. 8, no. 6, hlm. 3779–3783, Agu 2019, doi: 10.35940/ijeat.F9390.088619.
- [18] K. BA dan S. Charters, "Guidelines for performing Systematic Literature Reviews in Software Engineering," vol. 2, Jan 2007.
- [19] A. Dresch, D. P. Lacerda, dan J. A. V. Antunes, *Design Science Research: A Method for Science and Technology Advancement*. Springer Publishing Company, Incorporated, 2014.
- [20] P.Hemalatha, "Implementation of Data Mining Techniques for Weather Report Guidance for Ships Using Global Positioning System," dalam *International Journal Of Computational Engineering Research*, Mar 2013, vol. Vol. 3 Issue. 3.