

BAB 1. PENDAHULUAN

1.1 Latar Belakang

Pendidikan di Indonesia memiliki peran penting dalam pembangunan bangsa dan negara. Salah satu indikator kemajuan pendidikan adalah keberadaan dan kualitas karya tulis ilmiah. Karya ilmiah tidak hanya memperkaya khazanah ilmu pengetahuan, tetapi juga memberikan kontribusi nyata bagi masyarakat serta menjadi tolok ukur kualitas suatu lembaga pendidikan (Susanti, 2023). Dalam penulisannya, sitasi memegang peran penting karena memberikan landasan ilmiah, memperkuat argumen, serta menunjukkan akuntabilitas penulis terhadap sumber yang digunakan.

Karya ilmiah merupakan hasil pemikiran yang sistematis, berdasarkan pengamatan, penelitian, dan kajian mendalam, yang disusun dengan bahasa yang santun serta metodologi yang jelas (Nurfitriani et al., 2019). Untuk mempertanggungjawabkan setiap argumen atau pernyataan dalam karya ilmiah, penulis wajib menyertakan kutipan atau sitasi yang merujuk pada sumber yang relevan. Sitasi tersebut tidak hanya berfungsi sebagai penguat, tetapi juga sebagai penghubung antara karya ilmiah yang sedang dibuat dengan penelitian-penelitian sebelumnya.

Semakin berkembangnya teknologi informasi dapat didapatkan dengan mudah dan cepat, namun bersamaan dengan perkembangan teknologi muncul tantangan baru bagi peneliti dalam menilai kredibilitas dari sumber yang didapat (Apriandi, Boedi Maritasari, Maulida, & Karmila, 2024).

Dengan, meningkatnya jumlah publikasi ilmiah menyebabkan volume kalimat sitasi dalam satu dokumen pun semakin bertambah. Hal ini menyulitkan proses identifikasi fungsi sitasi secara manual karena membutuhkan waktu, ketelitian, dan pemahaman konteks. Di sinilah teknologi klasifikasi fungsi sitasi memiliki peran strategis. Seperti dijelaskan oleh (Teufel, Siddharthan, & Tidhar, 2006) pengenalan otomatis terhadap fungsi retorik sitasi memiliki berbagai

manfaat, seperti peningkatan akurasi ringkasan dokumen ilmiah, penghitungan metrik ilmiah yang lebih informatif, serta pengembangan sistem pencarian literatur yang lebih cerdas.

Dalam konteks tersebut, penelitian seperti SciCite yang dikembangkan oleh (Cohan, Ammar, Van, & Cady, 2019) menekankan pentingnya klasifikasi fungsi sitasi ke dalam kategori seperti Background, Method, dan Result. Klasifikasi ini dapat membantu pembaca dan sistem pencarian ilmiah memahami konteks serta kontribusi dari kutipan terhadap isi karya yang sedang dianalisis. Sayangnya, penelitian serupa dalam konteks Bahasa Indonesia masih sangat terbatas, padahal kebutuhan akan sistem otomatisasi klasifikasi fungsi sitasi juga tinggi di lingkungan akademik nasional.

Salah satu tantangan utama yang dihadapi penulis di Indonesia adalah kurangnya kemampuan dalam mencari dan mengelola sumber pustaka yang valid, serta menulis daftar pustaka secara sistematis (Lalu & Sri, 2019). Hal ini menyebabkan rendahnya kualitas sitasi yang digunakan, yang pada akhirnya dapat memengaruhi kualitas karya ilmiah secara keseluruhan. Oleh karena itu, dibutuhkan sebuah sistem berbasis teknologi yang mampu secara otomatis mengidentifikasi dan mengklasifikasikan fungsi dari kalimat sitasi dalam karya ilmiah berbahasa Indonesia.

Penelitian sebelumnya telah mencoba berbagai metode untuk mengklasifikasikan kalimat ilmiah. Misalnya, penggunaan Recurrent Neural Network (RNN) dan Word2Vec berhasil mencapai akurasi hingga 77,48% dalam membedakan fungsi sitasi ke dalam kelas Weak, Comparison, Point, dan Neutral (Firmansyah, Ilyas, & Kasyidi, 2020). Namun, pendekatan ini masih memiliki keterbatasan dalam memahami konteks semantik kalimat yang kompleks. Di sisi lain, metode klasifikasi tradisional seperti Support Vector Machine (SVM) menunjukkan akurasi lebih tinggi, yakni mencapai 98,5% ketika menggunakan ekstraksi fitur unigram (Awaliyah Zuraiyah, 2023), menjadikannya pendekatan yang potensial untuk diterapkan dalam konteks sitasi ilmiah berbahasa Indonesia.

Dengan demikian, penelitian ini bertujuan untuk menerapkan metode SVM dalam proses klasifikasi fungsi sitasi pada karya ilmiah berbahasa Indonesia. Dengan mengadaptasi pendekatan yang digunakan oleh SciCite dan menyesuaikannya dengan karakteristik Bahasa Indonesia, diharapkan sistem ini dapat membantu penulis akademik dalam menentukan sitasi yang relevan, kontekstual, dan berkualitas tinggi. Selain meningkatkan efisiensi proses penulisan, pendekatan ini juga berkontribusi terhadap peningkatan kualitas karya ilmiah secara menyeluruh, mengurangi risiko plagiarisme, dan memperkuat integritas ilmiah penulis (Iksan et al., 2022).

1.2 Rumusan Masalah

Berdasarkan masalah yang dijelaskan pada latar belakang, maka dapat dirumuskan masalah sebagai berikut:

- a. Bagaimana mengembangkan sistem berbasis web yang dapat memvalidasi kredibilitas sebuah journal berbahasa Indonesia?
- b. Bagaimana merancang dan mengembangkan sistem berbasis web yang mampu melakukan pencarian journal berdasarkan kata kunci?
- c. Bagaimana merancang dan mengembangkan sistem berbasis web yang memungkinkan pengguna melakukan input journal dan menghasilkan rekomendasi journal lain yang relevan secara otomatis?
- d. Bagaimana mengembangkan sistem berbasis web yang mampu mengklasifikasikan kalimat sitasi menggunakan metode SVM?
- e. Bagaimana akurasi model SVM menggunakan *confusion matrix*?

1.3 Tujuan

- a. Mengembangkan sistem berbasis web yang dapat memvalidasi kredibilitas jurnal berbahasa Indonesia, guna membantu pengguna dalam menilai keandalan sumber referensi ilmiah secara efisien dan objektif
- b. Merancang dan mengembangkan sistem pencarian jurnal berbasis web yang mampu menampilkan hasil berdasarkan kata kunci yang dimasukkan oleh pengguna, sehingga memudahkan akses terhadap referensi yang relevan.
- c. Mengembangkan sistem berbasis web yang memungkinkan pengguna melakukan input jurnal dan menghasilkan rekomendasi jurnal lain yang

relevan secara otomatis, untuk mendukung proses penelusuran literatur yang lebih efektif.

- d. Mengembangkan sistem klasifikasi kalimat sitasi berbasis web menggunakan metode Support Vector Machine (SVM), yang bertujuan untuk mengelompokkan jenis sitasi secara otomatis dan akurat dalam dokumen ilmiah.
- e. Mengukur akurasi SVM menggunakan *confusion matrix*.

1.4 Manfaat

Penelitian ini diharapkan memberikan manfaat bagi berbagai pihak, seperti akademisi, mahasiswa, peneliti, dan pengambil kebijakan. Adapun manfaat yang dapat diperoleh antara lain:

- a. Memberikan informasi mengenai peran dan fungsi kalimat dalam konteks sitasi, yang dapat dimanfaatkan oleh mahasiswa dan akademisi untuk menemukan sumber rujukan yang lebih tepat dan relevan dalam penulisan karya ilmiah
- b. Menyediakan sistem klasifikasi fungsi sitasi yang membantu dalam memahami konteks dan tujuan dari kutipan yang digunakan, sehingga pengguna dapat menyusun karya ilmiah dengan landasan teori yang lebih kuat dan terarah.
- c. Mendukung analisis bibliometrik yang lebih kaya dan informatif, dengan tidak hanya menampilkan jumlah sitasi, tetapi juga mengelompokkannya berdasarkan fungsi sitasi dalam dokumen, sehingga memberikan gambaran yang lebih mendalam terhadap pengaruh suatu karya ilmiah.
- d. Mempermudah pencarian referensi ilmiah berdasarkan kata kunci yang spesifik, sehingga pengguna dapat mengakses jurnal yang relevan dengan topik penelitian secara lebih efisien.
- e. Membantu dalam merekomendasikan jurnal terkait berdasarkan entri jurnal yang dimasukkan pengguna, sehingga mempercepat proses eksplorasi literatur dan mendukung kualitas studi literatur yang lebih komprehensif.

BAB 2. TINJAUAN PUSTAKA

2.1 Karya Peneliti Terdahulu

Peneliti terdahulu merujuk pada analisis karya tulis ilmiah dalam bentuk kalimat Tunggal dengan menggunakan beberapa metode berikut adalah beberapa penelitian:

- a. Klasifikasi Kalimat Ilmiah Menggunakan Recurrent Neural Network (Firmansyah, Ilyas, & Kasyidi, 2020)

Para akademisi tentunya ingin mempermudah dalam pekerjaan mereka terutama dalam mengklasifikasikan kalimat sitasi yang digunakan dalam karya tulis ilmiah. Dimulai dari penelitian ini masih belum ada model klasifikasi komputasi yang menggunakan metode *Recurrent Neural Network* (RNN). Dalam penelitian ini pengklasifikasian hanya berfokus pada satu kalimat ilmiah Tunggal dan tidak terpengaruh oleh kalimat sebelum dan juga sesudah.

Hal tersebut merupakan batasan masalah yang digunakan dalam penelitian ini. *Recurrent Neural Network* (RNN) merupakan jaringan saraf tiruan yang mampu digunakan dalam proses klasifikasi sesuai kelas yang diminta yaitu *weak*, *comparison*, *point*, dan *neutral*. Pada penelitian ini telah membandingkan empat metode optimasi yaitu Adam, SGD Adadelta, dan Adamax untuk menemukan tingkat pembelajaran terbaik dan yang cocok digunakan untuk klasifikasi kalimat ilmiah. dari 4 metode tersebut ternyata diperoleh Tingkat pembelajaran terbaik dengan menggunakan pengoptimalan SGD dengan nilai akurasi sebesar 71,48 % dan *loss* sebesar 0,71%. SGD dinilai cepat karena tidak menggunakan banyak *gradient*, *Desenct* sehingga *konvergen* lebih cepat. Selain itu SGD memilih sampel data secara acak dari data pelatihan dalam satu epoch training *dataset*, dengan cara yang *iterative*. Dari penelitian ini menunjukkan bahwa hasil akurasi skor *F-Measure* mencapai 39,5%.

b. Klasifikasi Kalimat Ilmiah Menggunakan 1D Convolutional Neural Networks (Firmansyah, Ilyas, & Kasyidi, 2020)

Kalimat-kalimat yang dituliskan dalam suatu karya tulis ilmiah memiliki banyak sekali ragamnya yang dipengaruhi dari hasil rujukan yang berbeda-beda. Hal tersebut tentunya jika dilakukan secara manual maka akan muncul kemungkinan terjadinya human error. Oleh sebab itu perlu adanya suatu model komputasi klasifikasi yang dapat membantu para Akademisi untuk mengklasifikasikan kalimat ilmiah ke dalam beberapa kelas dalam penelitian sebelumnya kelas dibagi menjadi 4 bagian yaitu “*weak*”, “*comparison*”, “*point*”, dan “*neutral*”.

Dalam proses klasifikasi pada penelitian ini menggunakan 2019 data berupa kalimat tunggal yang diperoleh dari makalah ilmiah komputasi dan telah dilabel dengan 4 kelas yang sudah disebutkan di penelitian sebelumnya. Dilakukan percobaan dengan menggunakan *CNN-Multichannel* dengan model optimasi Adam yang memiliki learning-rate 0,001 menghasilkan akurasi sebesar 74,51% dengan nilai loss sebesar 0,82. Lalu kemudian dilakukan percobaan menggunakan *CNN-Single channel* memperoleh akurasi sebesar 70,76 dan nilai loss sebesar 1,73. Lalu dilakukan serangkaian pengujian menunjukkan hasil terbaik dengan menggunakan *CNN-Multichannel* dengan skor *f-measure* sebesar 45,55%.

2.2 State Of The Art

Tabel 2. 1 *State of the art*

Tahun	Judul	Nama	Keterangan
2020	Klasifikasi Kalimat Ilmiah Menggunakan 1D <i>Convolutional Neural Networks</i>	Muhamad Rizal Firmansyah1 Ridwan Ilyas ,Fatan Kasyidi	Mengklasifikasi kalimat ilmiah menggunakan metode <i>CNN-Multichanel</i> kedalam 4 kelas yaitu <i>weak, comparison,point</i> dan <i>neutral</i> .
2020	Klasifikasi Kalimat Ilmiah Menggunakan <i>Recurrent Neural Network</i>	Muhamad Rizal Firmansyah ,Ridwan Ilyas,Fatan Kasyidi	Mengklasifikasi kalimat ilmiah menggunakan metode <i>Recurrent Neural Network</i> kedalam 4 kelas yaitu <i>weak, comparison,point</i> dan <i>neutral</i>
2024	Klasifikasi sitasi karya tulis ilmiah berbahasa Indonesia menggunakan metode <i>Support Vector Machine</i> (SVM)	Mohammad Tajut Zamzami	Penelitian ini akan mengklasifikasikan sitasi pada karya tulis ilmiah menggunakan metode <i>Support Vector Machine</i> (SVM) ke dalam 3 kelas yaitu <i>background, method, dan result</i>

2.3 Landasan Teori

2.3.1 Pendidikan

Pendidikan merupakan salah satu faktor penentu dalam meningkatkan sumber daya manusia (SDM) suatu bangsa. SDM yang berkualitas tentu dilandasi oleh nilai pendidikan yang ditanam dalam diri masyarakat. Selain itu pendidikan juga dapat membentuk karakter dan kemampuan individu sehingga bisa bersaing di pasaran dunia global (Laurensius Dihe Sangaa, 2023). Pendidikan dibagi menjadi 3 jenis yaitu (Siska, 2024).

a. Pendidikan Formal

Pendidikan formal adalah jenis pendidikan yang diberikan melalui sistem sekolah. Pendidikan formal terdiri dari beberapa tingkatan, yaitu pendidikan anak usia dini (PAUD), pendidikan dasar (SD/MI), pendidikan menengah pertama (SMP/MTs), pendidikan menengah atas (SMA/MA/SMK), dan pendidikan tinggi (perguruan tinggi, institut, atau universitas). Pendidikan formal dilaksanakan sesuai dengan kurikulum yang telah ditetapkan oleh pemerintah dan memiliki tujuan untuk memberikan pengetahuan umum, keterampilan, dan kemampuan berpikir kepada peserta didik (Siska, 2024).

b. Pendidikan Nonformal

Selain pendidikan formal, terdapat juga pendidikan nonformal. Pendidikan nonformal adalah pendidikan yang tidak mengikuti kurikulum formal yang ditetapkan oleh pemerintah. Pendidikan nonformal dirancang untuk memenuhi kebutuhan belajar masyarakat yang tidak memiliki akses atau kesempatan untuk mengikuti pendidikan formal. Jenis pendidikan nonformal meliputi kursus, pelatihan, bimbingan belajar, serta kegiatan pendidikan lainnya di luar lembaga pendidikan formal (Siska, 2024).

c. Pendidikan Informal

Selain pendidikan formal dan nonformal, terdapat juga pendidikan informal. Pendidikan informal adalah pendidikan yang tidak terstruktur dan tidak diatur oleh kurikulum formal. Pendidikan informal dapat terjadi melalui interaksi sosial sehari-hari, seperti melalui keluarga, teman, lingkungan, media massa, atau internet. Dalam pendidikan informal, pengetahuan dan keterampilan diperoleh secara tidak

sadar atau tidak langsung. Pendidikan informal memiliki peran penting dalam membentuk karakter dan pengetahuan individu (Siska, 2024).

2.3.2 Karya tulis ilmiah

Untuk berpikir kritis bisa dengan cara membuat sebuah karya tulis ilmiah (Susanti, 2023). Karya tulis ilmiah adalah hasil pemikiran ilmiah tentang disiplin ilmu tertentu yang disusun secara sistematis, logis, benar, holistik, dan bertanggung jawab dengan menggunakan bahasa yang baik dan benar (Heriyudananta, 2021). Beberapa ciri karya tulis ilmiah yang baik antara lain:

a. Mendalam/Tuntas.

Topik yang diangkat dalam karya ilmiah harus dianalisis secara mendalam hingga ke akar-akarnya. Sebagai penulis, kita sebaiknya tidak memilih topik yang terlalu luas. Sebagai contoh, kita dapat menggali lebih dalam tentang ‘Upaya Memerangi Korupsi di Indonesia (Abdul, 2017).

b. Objektif

Semua informasi yang disajikan dalam tulisan tersebut adalah akurat dan sesuai dengan data dan fakta yang diperoleh. Keobjektifan dalam karya ilmiah dapat dicapai dengan memiliki data literatur dan data lapangan yang memadai, serta dengan menghindari manipulasi data (Abdul, 2017).

c. Sistematis

Dalam konteks penulisan, artinya adalah bahwa uraian atau tulisan disusun dengan mengikuti pola tertentu sehingga terdapat urutan yang jelas dan kaitan yang berkelanjutan antara unsur-unsur dalam tulisan tersebut. Dengan demikian, pembaca dapat memahami hubungan antara setiap bagian tulisan secara berurutan dan terstruktur (Abdul, 2017).

d. Cermat

Sebagai penulis karya ilmiah, penting untuk bersikap cermat dalam setiap tahap penulisan. Ketelitian diperlukan dalam mengutip sumber, menyajikan data, dan menulis dengan ejaan yang benar. Kesalahan kecil dalam penulisan atau

pengutipan dapat berdampak besar, seperti kesalahan interpretasi atau tuduhan plagiarisme. Oleh karena itu, penulis perlu memastikan bahwa setiap kutipan telah disertai sumber yang valid dan setiap data disajikan sesuai dengan hasil yang diperoleh. Selain itu, peninjauan ulang terhadap struktur kalimat dan penggunaan tata bahasa yang baik juga menjadi bagian penting dari proses penulisan. Dengan bersikap cermat, penulis dapat menjaga kualitas dan kredibilitas karya ilmiahnya (Abdul, 2017).

e. Lugas

Lugas berarti menyampaikan informasi secara langsung, jelas, dan tanpa berbelit-belit. Dalam penulisan ilmiah, kelugasan penting agar pembaca mudah memahami isi tulisan tanpa kebingungan. Penulis perlu menghindari bahasa yang ambigu atau terlalu puitis, dan fokus pada inti persoalan. Gaya penulisan yang lugas mencerminkan sikap objektif dan profesional, serta membantu memperkuat argumen yang disampaikan (Abdul, 2017).

f. Tidak emosional

Dalam menulis karya tulis ilmiah, penulis harus menghindari melibatkan perasaan atau emosi pribadi. Penulisan harus bersifat objektif dan berdasarkan fakta atau data yang valid, bukan opini subjektif atau reaksi emosional. Hal ini penting agar argumen yang disampaikan dapat diterima secara ilmiah dan tidak bias. Dengan menjaga penulisan tetap netral dan logis, karya ilmiah menjadi lebih kredibel dan dapat dipertanggungjawabkan secara akademik (Abdul, 2017).

g. Logis

Dalam menulis karya tulis ilmiah, segala pernyataan dan argumen yang disampaikan harus didasarkan pada alasan yang logis, bukti yang kuat, serta data yang valid. Hal ini penting agar tulisan tersebut dapat dipertanggungjawabkan secara ilmiah dan memberikan kontribusi yang nyata dalam bidang yang dibahas. Dengan menggunakan dasar yang masuk akal, penulis dapat menghindari kesalahan interpretasi dan memastikan bahwa karya tulisnya memiliki kredibilitas serta kepercayaan dari pembaca dan komunitas akademik. (Abdul, 2017).

h. Bernas

Sebuah karya tulis ilmiah, tidak selalu harus panjang lebar, namun pastikan isinya padat dan relevan (Abdul, 2017). Kualitas lebih berarti daripada kuantitas (Kualitas vs Kuantitas: Perbedaan dan Perbandingan, 2024).

i. Jelas

Dalam menulis karya tulis ilmiah tentunya kita ingin karya kita mudah dipahami oleh orang lain oleh karena itu keterangan yang dikemukakan dapat mengungkap makna secara jernih sehingga mudah dipahami oleh pembaca (Abdul, 2017).

j. Terbuka

Dalam menulis karya tulis ilmiah tidak hanya terpaku dari pendapat peneliti saat itu sendiri namun juga ada pendapat baru yang tidak sesuai dengan pendapat yang ditulis di karya tulis ilmiah tersebut (Abdul, 2017).

k. Penggunaan Bahasa

Menggunakan bahasa baku, tepat, ringkas, dan jelas sangat penting dalam penulisan karya tulis ilmiah. Bahasa baku memastikan tulisan sesuai dengan kaidah tata bahasa yang berlaku sehingga mudah dipahami oleh pembaca. Pilihan kata yang tepat membantu menyampaikan maksud secara akurat tanpa menimbulkan kebingungan. Selain itu, kalimat yang ringkas dan jelas membuat informasi tersampaikan secara efektif tanpa bertele-tele, sehingga pembaca dapat menangkap inti pembahasan dengan mudah dan cepat (Abdul, 2017).

2.3.3 Natural Language Processing

Dalam bidang ilmu *computer* dan kecerdasan buatan, pemrosesan Bahasa Alami (*natural language processing*) atau NLP telah menjadi hal yang menarik. NLP memungkinkan *computer* untuk memahami dan mengolah teks manusia dalam berbagai Bahasa (Amien, 2023).

Di Indonesia dalam beberapa tahun terakhir penggunaan NLP telah berkembang pesat (Amien, 2023). NLP bertujuan untuk memahami konsep dan

maksud dari Bahasa manusia. Didalam NLP terdapat layanan untuk melakukan pemrosesan kata atau string (Amien, 2023).

2.3.4 Sitasi

Sitasi itu sendiri adalah kegiatan merujuk pada suatu artikel yang telah terbit pada suatu jurnal ilmiah dan lazim dilakukan oleh peneliti atau penulis untuk mendukung artikel yang diterbitkannya pada *journal* ilmiah (Saputro, 2022). Beberapa gaya sitasi yang umum digunakan adalah *American Psychological Association (APA)*, *Modern Language Association (MLA)*, *Medical Association (AMA)*, *Turabian*, *Chicago*, dan lain-lain (Imam, et al., 2019). Dalam penelitian sebelumnya penggunaan *class* dibagi menjadi 4 dimana hasil dari penyederhanaan *class* yang diambil dari Simone Teufel yang dapat dilihat pada Table 2.2.

Tabel 2. 2 Kelas Penelitian terdahulu

No	Kategori dari Teufel	Kategori yang digunakan	keterangan
1	Weak	Weak	Kelemahan dari
2	CoCo-		pendekatan yang dikutip/penelitian sebelumnya
3	CoCoR0	Comparison	Menggambarkan
4	CoCoGM		perbandingan atau
5	CoCoXY		kontras antara karya sendiridan lainnya
6	PBas	Point	Berfokus pada
7	PUse		sentimen positif
8	PUse		diekspresikan pada
9	PSim		kutipan, atau argumen/pernyataan penelitian sebelumnya digunakan dalam

No	Kategori dari Teufel	Kategori yang digunakan	keterangan
			penelitaian saat ini. seperti metode, data, algoritma, definisi
10	PMot	Neutral	Deskripsi netral dari karya yang dikutip, atau tidak cukup bukti tekstual untuk kategori di atas, atau fungsi kutipan tidak terdaftar
11	PSup		
12	Neut		

Sumber: (Firmansyah, Ilyas, & Kasyidi, 2020)

Sementara itu, dalam penelitian ini, klasifikasi tujuan sitasi menggunakan tiga kelas yang diambil dari *dataset SciCite* yang tertera pada Table 2.3.

Tabel 2. 3 Label yang digunakan

No	Kelas	Keterangan
1	Background	Kutipan menyatakan, menyebutkan, atau menunjukkan latar belakang informasi yang memberikan lebih banyak konteks tentang suatu masalah, konsep, pendekatan, topik, atau pentingnya masalah dilapangan.
2	Method	Kutipan yang penggunaannya Memanfaatkan suatu metode, alat, pendekatan, atau kumpulan data
3	Result	Perbandingan hasil/temuan makalah dengan hasil/temuan karya lain

Sumber: (Cohan, Ammar, Van, & Cady, 2019)

2.3.5 Multi Support Vector Machine (SVM)

Pada *support vector machine* (SVM) awalnya Klasifikasi tidak mendukung *multiclass* lalu semakin dikembangkan dan akhirnya bisa digunakan untuk melakukan klasifikasi *multiclass* (Octaviani, Wilandari, & Ispriyanti, 2014). Dalam

Klasifikasi *multiclass* pada SVM *hyperplane* yang terbentuk lebih dari satu. Metode yang cocok digunakan untuk pendekatan Klasifikasi *multiclass* adalah satu lawan semua (SLA). Metode SLA dapat digunakan untuk kasus Klasifikasi *k*-kelas, menentukan *k hyperplane* Dimana *k* adalah banyak kelas dan *p* adalah *hyperplane* dengan menggunakan fungsi *kernel Gaussian RBF* (Octaviani, Wilandari, & Ispriyanti, 2014).

2.3.6 Klasifikasi

Klasifikasi data adalah proses memprediksi kelas atau kategori berdasarkan fitur atau atributnya. Dalam dunia data *science* dan *machine learning*, klasifikasi merupakan bagian dari jenis *supervised learning*. Model yang digunakan dalam ranah klasifikasi di bidang machine learning adalah model yang dilatih untuk mengenali suatu *dataset* yang sudah diberi label (Firdausi, 2024).

2.3.7 Support Vector Machine

(Putra Mantovani, Sibaroni SSi, & Aditsania, 2016) menjelaskan bahwa SVM pertama kali diperkenalkan oleh Guyon Boser dan Vapnik, SVM adalah salah satu cara untuk melakukan klasifikasi dengan menggunakan Teknik regresi. Dengan kata lain SVM melakukan klasifikasi dengan cara menggunakan teori dari *machine learning* untuk memaksimalkan akurasi dengan menghindari *overfit* pada data secara otomatis. Svm memiliki 2 konsep dasar yaitu *margin hyperplane* dan juga *kernel* (Nugroho, Witarto, & Handoko, 2003).

Pada metode yang menerapkan *neural network* kebanyakan mencari *hyperplane* pemisah antar *class*, namun pada SVM berusaha menemukan *hyperplane* yang terbaik pada input space (Nugroho, Witarto, & Handoko, 2003). *Hyperplane* merupakan sebuah fungsi yang digunakan dalam memisahkan antar *class* atau kategori Irmanda, dkk. (2017) .

Sehingga jika dirumuskan maka untuk mencari nilai *hyperplane* yang terbaik adalah sebagai berikut dengan nilai *w* adalah *hyperplane* parameter yang ingin dicari *x* adalah data input SVM dan *b* adalah nilai bias atau parameter yang ingin dicari. Berikut merupakan persamaan untuk perhitungan hyperplane ditunjukkan.

$$w \cdot x + b = 0 \text{ ----- 2.1}$$

w = *weight vector*

x = *input vector*

b = *bias*

Berdasarkan persamaan 2.1 diatas dapat dilihat bahwa w merupakan nilai hyperplane yang dicari. Sedangkan x adalah data input SVM (x_1 = index kata, x_2 =bobot kata) dan b merupakan nilai bias atau parameter hyperplane yang dicari.

Adapun cara memaksimalkan margin adalah dengan cara mengubah x menjadi x_1 , y menjadi x_2 , a menjadi w_1 . B menjadi w_2 . Dengan persamaan seperti berikut

$$ax + by + c = 0 \text{ -----2.2}$$

- a = koefisien dari variable x , berhubungan dengan kemiringan garis terhadap sumbu x
- b = Koefisien dari variabel y , berhubungan dengan kemiringan garis terhadap sumbu y
- c = Konstanta atau bias (intersep), mengatur posisi garis terhadap titik asal.
- x,y = Koordinat pada bidang dua dimensi

2.3.8 Confusion Matrix

Untuk melakukan perhitungan akurasi pada konsep klasifikasi *confusion matrix* adalah salah satu metode yang cocok. *Confusion matrix dapat dilihat pada Tabel 2.4. (Rahman, Ilham Darmawidjadja, & Alamsah, 2017)*

Tabel 2. 4 Contoh Confusion Matrix

<i>Correct Classification</i>	<i>Classified as</i>	
	<i>Predicted “+”</i>	<i>Predicted “-”</i>
<i>Actual “+”</i>	<i>True Positives</i>	<i>False Negatives</i>
<i>Actual “-”</i>	<i>False positives</i>	<i>True Negatives</i>

Sumber : (Rahman, Ilham Darmawidjadja, & Alamsah, 2017)

Berdasarkan Tabel *confusion matrix* pada Tabel 2.4 dapat dijabarkan sebagai berikut:

- 1) *True Positives* (TP) adalah jumlah record data positif yang diklasifikasikan sebagai nilai positif (Rahman, Ilham Darmawidjadja, & Alamsah, 2017).
- 2) *False Positives* (FP) adalah jumlah record data negatif yang diklasifikasikan sebagai nilai positif (Rahman, Ilham Darmawidjadja, & Alamsah, 2017).
- 3) *False Negatives* (FN) adalah jumlah record data positif yang diklasifikasikan sebagai nilai negative (Rahman, Ilham Darmawidjadja, & Alamsah, 2017).
- 4) *True Negatives* (TN) adalah jumlah record data negatif yang diklasifikasikan sebagai nilai *negative* (Rahman, Ilham Darmawidjadja, & Alamsah, 2017)

Dari perhitungan *confusion matrix* akan menghasilkan nilai sebagai berikut:

- 1) *Accuracy, presentase* jumlah record data yang diklasifikasikan (prediksi) secara benar oleh algoritma (Rahman, Ilham Darmawidjadja, & Alamsah, 2017).

$$Accuracy = \frac{TP + TN}{TOTAL DATA} \text{-----} 2.3$$

- 2) Presisi merupakan proporsi prediksi benar (True Positive) terhadap seluruh prediksi positif yang dibuat oleh model untuk setiap kelas (Hafidz & Yanti Liliana, 2021).

$$Precision = \frac{TP}{TP + FP} \text{-----} 2.4$$

- 3) *Recall* atau yang juga dikenal sebagai *True Positive Rate*, adalah rasio dari seluruh data positif yang berhasil diidentifikasi dengan benar oleh model (Hafidz & Yanti Liliana, 2021).

$$Recall = \frac{TP}{TP + FN} \text{-----} 2.4$$

2.3.9 User Acceptance Testing

User Acceptance Testing (UAT) adalah proses pengujian yang dilakukan oleh pengguna akhir yang secara langsung menggunakan sistem untuk memastikan bahwa seluruh fungsi berjalan sesuai dengan tujuan dan kebutuhan yang telah ditentukan. Pengujian ini menggunakan pendekatan black box testing, yaitu dengan memeriksa sistem berdasarkan spesifikasinya tanpa melihat struktur internal dari perangkat lunak (Wahyudi & Alameka, 2023).

UAT berfokus pada evaluasi sistem oleh pengguna untuk memastikan bahwa aplikasi benar-benar memberikan manfaat dan dukungan yang dibutuhkan dalam aktivitas mereka. Tahapan ini menjadi bukti bahwa sistem siap digunakan secara operasional karena telah terbukti mampu memenuhi kebutuhan riil pengguna, bukan sekadar kesesuaian terhadap spesifikasi teknis yang dirancang (Wahyudi & Alameka, 2023).

2.3.10 Flowchart

Flowchart merupakan representasi sistematis dari proses dan logika yang digunakan dalam pengelolaan informasi. Diagram ini menggambarkan langkah-langkah serta urutan prosedur dari suatu program secara grafis. *Flowchart* membantu analis dan programmer dalam memecah permasalahan menjadi bagian-bagian yang lebih kecil, sekaligus mempermudah dalam mengevaluasi berbagai alternatif solusi yang mungkin digunakan dalam sistem. Adapun *system flowchart* menggambarkan alur proses dalam suatu sistem, termasuk media input dan output yang digunakan, serta jenis media penyimpanan yang terlibat dalam proses pengolahan data (Bagus A P & Handrianus Pranatawijaya, 2018).

2.4 Kajian Teknologi

2.4.1 Perangkat Lunak (Software)

Perangkat lunak adalah program yang berisi Kumpulan instruksi untuk melakukan proses pengolahan data. Perangkat lunak menjadi penghubung antara manusia dengan perangkat keras yaitu komputer. Ada sebuah penerjemah yang bertugas menerjemahkan Bahasa manusia ke Bahasa yang digunakan oleh komputer sehingga komputer dapat memberikan hasil yang diinginkan (Sudarso, Fakultas Ekonomi, Bhayangkara, & Raya, 2021).

2.4.2 Website

Sebuah website adalah kumpulan halaman web beserta file-file pendukungnya, seperti gambar, video, dan berbagai file digital lainnya. Semua ini disimpan pada sebuah server web yang biasanya dapat diakses melalui internet. Dengan kata lain, website terdiri dari sejumlah folder dan file yang mengandung berbagai perintah dan fungsi, termasuk fungsi tampilan dan fungsi untuk mengelola penyimpanan data (Suhartini, Muhamad Sadali, & Yupi Kuspani Putra, 2020).

2.4.3 Python

Python adalah salah satu bahasa pemrograman yang relatif baru dan populer saat ini. Dalam Python, kita dapat membuat program dengan lebih sederhana dan singkat. Setiap program yang kita buat pasti memerlukan input dan output. Meskipun metode penginputannya terlihat mudah, tetapi sebenarnya masih banyak pemula yang mengalami kesulitan dalam membuat program menggunakan Python. Dengan adanya pembahasan ini, diharapkan dapat membantu pemula yang sedang belajar bahasa pemrograman.

2.4.4 PHP

PHP adalah bahasa pemrograman yang secara khusus dikembangkan untuk keperluan pengembangan web. Bahasa ini digunakan sebagai alat untuk membuat halaman web yang bersifat dinamis. Awalnya, PHP merupakan singkatan dari *Personal Home Page* dan pertama kali dikembangkan oleh Rasmus Lerdorf pada tahun 1995. Pada saat itu, PHP masih dikenal dengan nama *FI Form Interpreter*, yang berupa kumpulan skrip sederhana untuk memproses data dari formulir web. Seiring perkembangan waktu, PHP kini dikenal sebagai singkatan dari *PHP: Hypertext Preprocessor*, sebuah akronim rekursif di mana bagian awal kepanjangan mencakup singkatannya sendiri (Lutfi, 2017).

2.4.5 Laravel

Laravel merupakan *framework* PHP *open-source* yang tersedia secara gratis dan dikembangkan oleh Taylor Otwell. *Framework* ini dirancang untuk mempermudah pengembangan aplikasi web dengan menerapkan pola arsitektur *model-view-controller* (MVC). Laravel menawarkan berbagai fitur, seperti sistem modular yang dapat dikelola dengan baik, pendekatan baru dalam mengakses

database relasional, serta menyediakan berbagai utilitas untuk memudahkan proses deployment dan pemeliharaan aplikasi (Sahrul, et al., 2016).